Investigating genetic determinants of ankylosing spondylitis

Thesis

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.21954/ou.ro.0000e81c

INVESTIGATING GENETIC DETERMINANTS OF ANKYLOSING SPONDYLITIS

A thesis submitted in partial fulfilment of the requirements of the Open University for the degree of Doctor of Philosophy

Elisa Johanna Jaakkola

Licentiate in Medicine

Wellcome Trust Centre for Human Genetics,

Sponsoring establishment:

Weatherall Institute of Molecular Medicine,

Oxford

December 2003

Elisa Johanna Jaakkola, Licentiate in Medicine
The Open University
Sponsoring establishment: Weatherall Institute of Molecular Medicine, Oxford
Thesis submitted for the degree of Doctor of Philosophy, 2003

INVESTIGATING GENETIC DETERMINANTS OF ANKYLOSING SPONDYLITIS

Ankylosing spondylitis (AS) is a common form of inflammatory arthritis, occurring in approximately 0.1% of the British and Finnish populations. The genetic contribution to the disease susceptibility accounts for >90% of the population variance, and the disease severity is also predominantly genetically determined. *HLA-B27* has been strongly associated with the disease world-wide, but although it is almost essential for the inheritance of AS, the attributable risk of *HLA-B27* has been estimated to be 16-50%, leaving a large component of the genetic variance to be discovered.

The influence of a positional and biological candidate gene, *transforming growth factor β1 (TGFB1)*, was investigated in a Finnish and British population. A weak association between the rare *TGFB1* 1627 T allele and susceptibility to, and age of symptom onset of, AS was noted. However the lack of association of *TGFB1* promoter polymorphisms, which are in LD with the *TGFB1* 1627 alleles, with AS in families with positive non-parametric linkage scores at the locus indicates that these polymorphisms do not explain the observed linkage of disease susceptibility to chromosome 19.

A novel high throughput *HLA-DRB1* genotyping method based on multiplex primer extension reactions was developed. This method allows rapid and cost-effective screening of a large number of samples.

The effect of *HLA* genes and haplotypes in susceptibility to, and severity of, AS was investigated in the Finnish population. An overrepresentation of *HLA-B27* homozygotes was noted in cases with AS compared with the expected number of *HLA-B27* homozygotes under Hardy-Weinberg equilibrium. Significant associations between both *HLA-B27* and *HLA-DRB1*08* and a younger age of symptom onset was noted suggesting that genes within the MHC are involved in determining the age of symptom onset. A haplotype-based case-control study noted no association between *HLA-DRB1-B27* haplotypes and AS susceptibility, but this study was underpowered.

The effect of *CYP2D6*4* poor metaboliser allele was investigated in the Finnish AS families. No association was noted, but due to lack of power this study could not exclude a true positive association with the disease.

In summary, a novel *HLA-DRB1* genotyping technique was developed to enable the assessment of the contribution of *HLA-DRB1* genes in a large number of AS samples. *HLA-B27* homozygosity was increased in cases with AS compared with the expected number of *HLA-B27* homozygotes. *HLA-B27* and *HLA-DRB1*08* alleles were significantly associated with a younger age of symptom onset in AS. *TGFB1* gene polymorphisms do not appear to have a major impact in AS.

PUBLICATIONS ARISING FROM THIS THESIS

Jaakkola E, Crane AM, Laiho K, Herzberg I, Sims AM, Bradbury L, Calin A, Brophy S, Kauppi M, Kaarela K, Wordsworth BP, Tuomilehto J, Brown MA (2004) The effect of transforming growth factor beta-1 gene polymorphisms in ankylosing spondylitis. Rheumatology (Oxford) 43:32-8

Jaakkola E, Herzberg I, Crane AM, Pointon J, Laiho K, Kauppi M, Kaarela K, Wordsworth BP, Tuomilehto J, Brown MA (2004) A novel HLA-DRB1 genotyping method based on multiplex primer extension reactions. Tissue Antigens (in press)

Jaakkola E, Herzberg I, Pointon J, Sims AM, Barnardo M, Laiho K, Kauppi M, Kaarela K, Tuomilehto-Wolf E, Tuomilehto J, Ilonen J, Wordsworth BP, Brown MA (2004) The effect of HLA genes in susceptibility to and severity of ankylosing spondylitis in a Finnish population. (manuscript in process)

OTHER RELATED PUBLICATIONS

Goedecke V, Crane AM, **Jaakkola E**, Kaluza W, Laiho K, Weeks DE, Wilson J, Kauppi M, Kaarela K, Tuomilehto J, Wordsworth BP, Brown MA (2003) Interleukin 10 polymorphisms in ankylosing spondylitis. Genes Immun 4:74-6

ATTRIBUTIONS

The Finnish families with AS were recruited by Dr Kari Laiho, Dr Markku Kauppi, Dr Kalevi Kaarela and Arja Lyytikäinen from the Rheumatism Foundation Hospital in Heinola, Finland. Ibi Herzberg, Anne-Marie Sims, Owen Beynon, Andrew Timms and Adrian Edwards extracted approximately half of the DNA of the Finnish families presented in this thesis, the rest of the DNA was extracted by myself. The British families with AS were recruited, and DNA extracted, by Dr M. Brown and colleagues.

Chapter 3: The microsatellite genotyping for the *TGFB1* study was carried out as part of the fine mapping of the whole genome screen by Sarah Edwards and Dr Alison Crane. Ibi Herzberg performed half of the SNP genotyping of *TGFB1* promoter region in the British affected sibling pair data set.

Chapters 4 and 5: Ibi Herzberg performed the PCR reactions for HLA-DRB1*02 subtyping and some of the *HLA-B27* genotyping using the PCR-SSP. The control samples for the HLA case-control study were typed by Eva Tuomilehto-Wolf and colleagues in Finland. The *HLA-B27* homozygous/heterozygous genotyping method based on PCR-SSP was developed by Dr Martin Barnardo and Juliet Agudelo from the Tissue Typing laboratory, Churchill hospital, Oxford. The sequence based method for *HLA-DRB1* was developed by Dr Steven Laval. Prof Daniel E Weeks developed the statistical method to calculate the expected number of homozygote individuals under HWE. The rest of the laboratory work, genotyping and statistical analysis was carried out by myself.

ACKNOWLEDGEMENTS

This work was carried out in the Wellcome Trust Centre for Human Genetics, Oxford during the years 2000-2003. I would like to thank all the people who have made it possible for me to undertake this project.

I would like to thank Dr Matthew Brown for allowing me to work on this project and for his supervision and guidance. I would also like to thank my second supervisor, Dr Paul Bowness, for the constructive advice provided when reading my thesis.

I would also like thank Prof Paul Wordsworth for his support throughout and for finding funding for my last year. Particular thanks to Dr Jennifer Pointon for reading through my thesis and useful comments. Thank you to Dr Alison Crane for her advice. Special thanks to Ibi Herzberg for giving excellent technical assistance and being a true friend. Thanks to Dr Anita Milicic for helping me a lot at the beginning of the project and also thanks to all the other members and ex-members of the Brown group for their help. Thank you to the members of the core facilities for preparing gels for sequencing and running the plates for the Snapshot.

Thank you to our collaborators in Finland, Prof Jaakko Tuomilehto, Dr Kari Laiho, Arja Lyytikäinen, Dr Markku Kauppi and Dr Kalevi Kaarela, for recruiting the Finnish families and collecting phenotype data. Thanks to all the families and cases who have participated in this study.

Most of all, I would like to thank my husband Sandy for putting up with late nights, the financial support, encouragement and invaluable help with computer matters. Loving thoughts to our son Hannu, whose birth has brought so much joy to my life. Special thanks to my parents for their support and to my husband's parents for their support and help in particular with the childcare.

LIST OF ABBREVIATIONS

µl	microlitre
µM	micromolar
3'	3 primer end of the sequence
5'	5 prime end of the sequence
A	adenine
AAU	acute anterior uveitis
AFBAC	affected family based controls
AIMS	arthritis impact measurement scales
ANCOVA	analysis of covariance
ANOVA	analysis of variance
ARMS	amplification refractory mutation system
AS	ankylosing spondylitis
BASDAI	Bath AS disease activity index
BASFI	Bath AS functional index
BAS-G	Bath AS global score
BiP	immunoglobulin heavy chain binding protein
bp	base pair
c.	cDNA
C	cytosine
CARD	caspase recruitment domain
CD/CV	common disease / common variant hypothesis
CD/RA	common disease / rare allele hypothesis
CI	confidence interval
cM	centiMorgans
CT	computed tomography
CTL	cytotoxic T lymphocyte
CYP2D6	cytochrome P450 2D6, debrisoquine 4-hydroxylase
ddNTP	dideoxynucleotide triphosphates
DNA	deoxy-ribonucleic acid
dNTP	deoxynucleotide triphosphate
dsDNA	double-stranded DNA
DZ	dizygotic
E	entropy
EDTA	ethylenediamine tetraacetic acid
ELISA	enzyme linked immunosorbent assay
ER	endoplasmic reticulum
ethidium bromide	2,7-diamino-10-ethyl-9-phenylphenanthridinium bromide
Exo I	exonuclease I
G	guanine
GAS	genetic analysis system program
GOLD	graphical overview of linkage disequilibrium
GRR, γ	genotype relative risk
GuHCl	guanidinium hydrochloride
HapMap	haplotype mapping

HHRR	haplotype-based relative risk
HLA	Human Leukocyte Antigen
HRR	haplotype relative risk
HSP	heat shock protein
htSNP	haplotype tagging SNPs
HWE	Hardy-Weinberg equilibrium
IBD	inflammatory bowel disease
ibd	identical by descent
IDDM	insulin-dependent diabetes mellitus
IFNγ	interferon γ
IL	interleukin
IL-1Ra	IL-1 receptor antagonist
Ig	immunoglobulin
IVS	intervening sequence
LAP	latency-associated polypeptide
LTBP	latent TGFB-binding protein
LD	linkage disequilibrium
LMP	large-molecular-weight proteasome
LOD	logarithm of the odds
MALDI-TOF	matrix-assisted laser desorption-ionization-time-of-flight
Mb	megabase
MEGA2	manipulation environment for genetic analyses
MERLIN	multipoint engine for rapid likelihood inference
MHC	Major Histocompatibility Complex
MICA	MHC class I chain related A
min	minutes
mM	millimolar
MRI	magnetic resonance imaging
mRNA	messenger RNA
MZ	monozygotic
NF-κB	nuclear factor kappa B
ng	nanogram
$(NH_4)_2SO_4$	ammonium sulphate
NK	natural killer
NPL	non-parametric linkage
NS	not significant
°C	degrees Centigrade
OD	optical density
OMIM	online Mendelian inheritance in man (www.ncbi.nlm.nih.gov/Omim/)
OPLL	ossification of the posterior longitudinal ligament
OR	odds ratio
PCR	polymerase chain reaction
PCR-RFLP	PCR restriction fragment length polymorphism
PCR-SBT	PCR sequencing based typing
PCR-SSO	PCR sequence-specific oligonucleotides
PCR-SSP	PCR sequence specific primers
PDGF	platelet derived growth factor
pip	percent identity plot
polyA	polyadenylation signal

QTDT	quantitative transmission disequilibrium test
RA	rheumatoid arthritis
rcf	relative centrifugal force or g-force
ReA	reactive arthritis
RPE	relative predispositional effect
rpm	revolution per minute
RR	relative risk
s	seconds
SAP	shrimp alkaline phosphatase
S.D.	standard deviation
SDS	sodium dodecyl sulphate
SNP	single nucleotide polymorphism
SpA	spondyloarthropathy
STR	short tandem repeat
T	thymine
TAP	transporter associated with antigen processing
TBE	Tris-Borate-EDTA buffer
TDT	transmission disequilibrium test
TE	Tris-EDTA buffer
TGFB / TGF-β	transforming growth factor β gene/protein
Th	T helper cell
™	Registered Trademark
TNF-α	tumour necrosis factor alpha
UV	ultraviolet
θ	recombination fraction
λs	sibling relative risk ratio
T_m	melting temperature

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 - INTRODUCTION

1.1 ANKYLOSING SPONDYLITIS

1.1.1 Clinical characteristics

Ankylosing spondylitis (AS) (OMIM reference number 106300) is a chronic, systemic, inflammatory rheumatic disease that predominantly involves the axial skeleton including the sacroiliac joints and the spine. The presence of sacroiliitis is its hallmark. The other axial sites commonly affected include apophyseal, discovertebral, costovertebral and costotransverse joints of the spine, and the paravertebral ligamentous structures. Hip and shoulder joints and less commonly peripheral joints may be involved. The presence of hip involvement correlates positively with a young age of onset and disease severity (Brophy and Calin 2001). Peripheral arthritis typically affects large synovial joints and occurs in approximately 30% of patients. Extra-articular manifestations include acute anterior uveitis (AAU), aortic valve disease and enteric mucosal inflammatory lesions (Calin and Taurog 1998).

AS belongs to the group of related inflammatory joint diseases termed spondyloarthropathies (SpA). They share characteristic clinical features and an association with the human MHC class I molecule HLA-B27. Other SpA include reactive arthritis, psoriatic arthritis, arthritis associated with inflammatory bowel disease (IBD) and undifferentiated SpA (Dougados et al. 1991). In most cases, primary AS is not associated with any other disorder, but secondary AS occurs in association with reactive arthritis, psoriasis, ulcerative colitis or Crohn's disease. The strength of association with HLA-B27 varies among different SpA, ranging from 50-60% in psoriatic and enteropathic spondylitis, to 60-80% in undifferentiated SpA and reactive arthritis, and to over 90% in primary AS. Reactive arthritis is triggered by bacteria infecting the urogenital tract (*Chlamydia trachomatis*), the gut (*Yersinia, Salmonella, Shigella and Campylobacter*) and possibly also the upper respiratory tract (*Chlamydia pneumoniae*) (Ramos and De Castro 2002). The involvement of microbes has often been suggested in AS, but no triggering pathogenic agents have been identified.

AAU is the most frequent extra-articular manifestation in patients with AS. Up to 30% of AS patients manifest AAU and it is more common in *HLA-B27* positive than negative patients. The uveitis associated with AS is typically acute in onset, unilateral, frequently recurrent and spares the choroid and the retina (Rosenbaum 1989). This recurrent unilateral AAU can also occur as an isolated disease entity without joint involvement but with a strong association with *HLA-B27* antigen (Brewerton et al. 1973a).

AS and IBD are clinically associated diseases, with studies indicating that 10% of cases of IBD are complicated by AS (de Vlam et al. 2000). Clinically proven IBD is present in 4% of AS patients (Edmunds et al. 1991). Clinically silent enteric mucosal inflammatory lesions in the terminal ileum and colon have been detected by endoscopy in 44–69% of patients with AS (Leirisalo-Repo et al. 1994; Mielants et al. 1995). The role of these mucosal changes in the pathogenesis of AS is unknown, but it has been suggested that abnormal mucosal permeability in the ileum may allow pathogenic antigens to enter the circulation and trigger AS. Clinical remission of SpA is associated with normal gut histology, whereas active locomotor inflammation was usually associated with the presence of gut inflammation (Mielants et al. 1995). The gut inflammation is clinically and histologically closely related with gut inflammation seen in Crohn's disease (Baeten et al. 2002). Anti-*Saccharomyces cerevisiae* IgA antibodies are considered to be important marker for Crohn's disease. A recent study noted significantly higher levels of anti-*Saccharomyces cerevisiae* IgA antibodies in patients with AS than in healthy controls (Hoffman et al. 2003). This may indicate a possible common pathogenic mechanism underlying these two diseases, or IgA antibodies may be a marker of small bowel ulceration. Bowel inflammation is clinically associated with peripheral joint inflammation (Baeten et al. 2002).

Other nonarticular structures, such as the aortic valve, lungs and kidneys can be affected, but these complications are rare. Three types of heart involvement in AS have been noted: aortitis and aortic insufficiency, conduction disturbances of the atrioventricular node and myocardial involvement with a possible compromise of left ventricular function (Lautermann and Braun 2002). Aortitis of ascending aorta can result in the dilatation of the aortic ring and aortic valve incompetence. A slowly progressive apical pulmonary fibrosis is a rare complication of AS (Rosenow et al. 1977), as are amyloidosis and uraemia. IgA nephropathy is also a reported complication of AS, which is associated with raised serum IgA levels.

1.1.2 Diagnostic criteria of AS

AS patients in our studies were diagnosed using the Modified New York Criteria (1984) (van der Linden et al. 1984a) as follows.

1. Clinical criteria

    a) Low back pain of at least 3 months' duration improved by exercise and not relieved by rest

    b) Limitation of lumbar spine in sagittal and frontal planes

    c) Decreased chest expansion relative to normal values for age and sex

2. Radiologic criterion

    Bilateral sacroiliitis grade 2 to 4 or unilateral sacroiliitis grade 3 to 4 on X-ray.

    The following scoring of the radiographs was used: grade 0 = normal; grade 1 = suspicious changes; grade 2 = minimum abnormality (small localised areas with erosion or sclerosis, without alteration in the joint width); grade 3 = unequivocal abnormality (moderate or advanced sacroiliitis with erosions, evidence of sclerosis, widening, narrowing or partial ankylosis); grade 4 = severe abnormality (total ankylosis).

Definite AS if the radiologic criterion is associated with at least 1 clinical criterion.

Imaging in AS. The onset of AS is insidious, progressing to radiological sacroiliitis over several years. The development of radiographic bone damage may be slow, taking up to at least a decade to manifest (Mau et al. 1988). Radiography lacks the sensitivity to show small focal changes in skeletal calcium content in large joints. Figure 1.1 illustrates typical radiological features of AS.

Recent data suggest that magnetic resonance imaging (MRI) of the sacroiliac (SI) joints can be used to identify sacroiliitis earlier than plain radiography (Oostveen et al. 1999). Early sacroiliitis can be demonstrated by dynamic MRI in SpA patients in whom abnormalities are not revealed by conventional radiography (Braun et al. 1994). However, the specificity of the MRI is unknown and therefore it has not yet been included in diagnostic criteria.

Figure 1.1 Clinical and radiological features of AS.

a) Early sacroiliitis. The arrows point to the sacroiliac joints, which are shown in Ferguson projection, with the patient supine and the tube angled caudocephalad approximately 30°. This allows a view of the entire length of the sacroiliac joints symmetrically. Subchondral bone resorption and irregularity of the sacroiliac joint spaces give rise to the apparent pseudowidening. Increased sclerosis is present around the sacroiliac joints.

b) Advanced sacroiliitis. The sacroiliac joints are nearly completely obliterated. Bony trabeculae cross the residual sacroiliac joint space. Bony ankylosis results from endochondral ossification. No gross sclerosis is seen in this radiograph. A moderate degree of osteopenia is present.

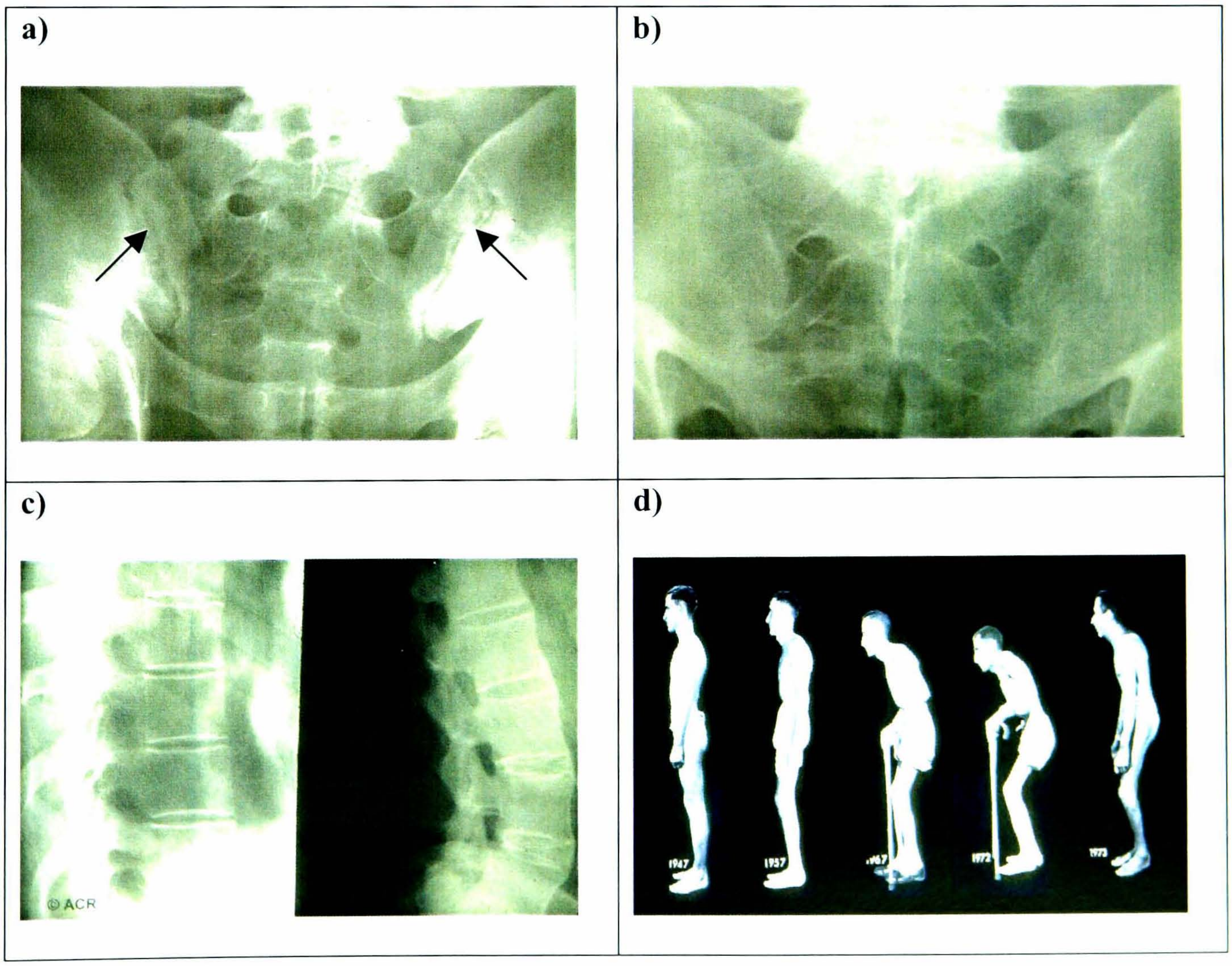
c) Thoracic (left) and lumbar (right) spine show squaring of the vertebral bodies, osteopenia and ossification of the anterior longitudinal ligaments.

Left: A lateral projection of the lower thoracic spine shows osteopenia. The intervertebral spaces are preserved. There are loss of the anterior concavity of the vertebral bodies and straightening of the anterior margins. The squaring of the vertebral bodies is caused by anterior spondylitis with resorption at the enthesis. Squaring is pathologic when seen in the lumbar spine, but it may be normal in the thoracic spine.

Right: A lateral view of the lumbar spine shows moderate osteopenia and ossification of the anterior longitudinal ligaments. The intervertebral spaces are preserved, and the configuration of the vertebral bodies is relatively normal. In this case, the lumbar lordosis is maintained.

d) The progression of deformities of a patient with AS over a period of 26 years. The clinical changes include thoracic kyphosis and the loss of lumbar lordosis and subsequent development of increasing flexion contractures of the hips and knees along with progressive ankylosis of the spine. In 1973, bilateral total hip arthroplasties were performed, which improved the patient's posture.

The figures were obtained from a CD-ROM compiled by the American College of Rheumatology.



1.1.3 Epidemiology

A Finnish study has reported an AS prevalence of 0.13% (Kaipiainen-Seppanen et al. 1997), similarly a Dutch study reported an AS prevalence of 0.1–0.2% (van der Linden et al. 1984b). In contrast studies from Norway and Germany reported AS prevalences of 1.1–1.4% and 0.9%, respectively (Gran et al. 1985; Braun et al. 1998). The differences between these studies are likely to be attributable to the diagnostic criteria used, differences in the population frequency of *HLA-B27*, response rate of the study or other factors related to study methodology.

The symptoms usually begin in late adolescence or early adulthood and the average age of symptom onset is 25 years (Feldtkeller et al. 2003). The disease is more common in males, with the male to female ratio about 2.5:1 to 5:1. Clinical manifestations appear to be similar in men and women, whereas radiological features appear to be more frequent and severe in males (Gran and Ostensen 1998).

1.1.4 Pathology of AS

Classically, AS begins at the sacroiliac joints and then ascends the spinal column by involving the small joints of the posterior elements of the spine. Sacroiliitis, enthesitis and osteitis are characteristic primary spinal lesions in AS. The inflammation affects the synovium, articular cartilage and subchondral bones of involved joints and can also involve juxta-articular ligamentous structures. Plasma cells, lymphocytes and polymorphonuclear cells infiltrate the sites of enthesitis. Inflammatory infiltrates and oedema may be present at the adjacent bone-marrow space. The inflammatory process results in initial bone and joint erosion. The inflammation frequently results in gradual fibrosis and ossification, leading to subsequent syndesmophyte formation (with eventual bony bridging) and ankylosis (Calin and Taurog 1998) (see Figure 1.1).

Enthesitis is an important but not the sole pathologic lesion in AS (McGonagle et al. 1998a; Francois et al. 2001). An enthesis is the region where a tendon, ligament or joint capsule attaches to bone. Tendon enthesis can be classed as fibrous or fibrocartilaginous according to the tissue present at the skeletal attachment site. Fibrocartilaginous entheses consist of uncalcified fibrocartilage, a proteoglycan-rich matrix and calcified fibrocartilage. Fibrous entheses have no cartilage matrix and consist purely of dense fibrous connective tissue, where the characteristic cell type is the fibroblast (McGonagle et al. 2003). It has been suggested that AS mainly affects the fibrocartilaginous enthesis (Maksymowych 2000). Fat-suppressed MRI scans of knee joints of SpA patients demonstrated prominent entheseal abnormalities which were not present in knee joints of patients with rheumatoid arthritis (RA). This supports the central role of enthesis in the pathology of SpA (McGonagle et al. 1998b).

Enthesitis is closely associated with adjacent osteitis. The pathogenesis of osteitis relates to bone microfractures or an abnormal response to bone stress adjacent to the insertions. The mechanism of osteitis is poorly understood, but tumor necrosis factor α (TNF-α) and interleukin (IL) 1 contribute to osteoclast activation and may play important roles (Braun et al. 1995; McGonagle et al. 2003). A recent study investigating tissue specimens of enthesis of SpA patients noted that oedema and an inflammatory infiltrate were predominant in the bone marrow component of the enthesis. The predominant T cells were CD8+ which suggests that these cells may have a key role in local inflammation in chronic SpA (Laloux et al. 2001).

Sacroiliitis is a hallmark of AS and the presence of radiologically confirmed sacroiliitis is included in the diagnostic criteria of the disease (van der Linden et al. 1984a). A recent systematic histopathological study of sacroiliitis which included 12 cases of AS and 22 control cases demonstrated that AS pathology represents more than enthesitis. In AS synovitis, myxoid bone marrow, cartilage destruction, enthesitis, new bone formation and bony ankylosis were significantly more frequent than in control cases. Mild but destructive synovitis and myxoid subchondral bone marrow were the earliest changes identified in AS. These changes gave rise to granulation tissue, which may be more important than enthesitis to explain the widespread destruction of cartilage and bone. A proliferative process of cartilage metaplasia and endochondral ossification contributes to the final ankylosis (Francois et al. 2000). Another study investigated computed tomography (CT)-assisted biopsy specimens of sacroiliac joints using immunohistology and *in situ* hydridisation. T cells, macrophages and various cytokines, especially TNF-α and transforming growth factor β2 (TGF-β2) were observed in infiltrates (Braun et al. 1995).

Bone sclerosis, new bone formation and syndesmophyte formation in addition to bone destruction are characteristic in AS. These bone changes result in joint ankylosis and cause disability in these patients. At the end stage of disease, the spine may become fused anteriorly resulting in the characteristic bamboo spine appearance in the radiograph. The unburdened vertebral bodies become 'square' and osteoporotic. Radiographs of the spine often reveal some degree of osteopenia and the risk of fracturing is increased. Radiographic and MRI studies in early AS suggest that MRI bone changes predate radiographic abnormalities suggesting that enthesitis and osteitis are primary and bone changes such as sclerosis or the formation of syndesmophytes are secondary (Oostveen et al. 1999).

Osteoporosis may also be a primary pathological event. Several studies have reported that patients with early AS demonstrate a significantly lower bone mineral density both in the lumbar spine and in the femoral neck than controls (Will et al. 1989; El Maghraoui et al. 1999). Possible mechanisms explaining the early osteoporosis in AS include genetic predisposition, the production of cytokines, such as transforming growth factor β1 (TGF-β1), IL-1 and TNF-α, involved in bone remodelling (Centrella et al. 1991), and impaired calcium and vitamin D absorption related to chronic inflammatory intestinal lesions (Mielants et al. 1995). Serum levels of vitamin D have been reported to correlate negatively with disease activity in AS (Lange et al. 2001). Decreased vitamin D may contribute to a negative calcium balance and inhibition of bone formation. In the late stage of the disease, bony ankylosis is responsible for the osteoporosis.

The mechanisms leading to immune activation in AS are poorly defined. A multi-hit model, in which biomechanical stressing, tissue microtrauma and microbial factors at the entheses induce the expression of pro-inflammatory cytokines and incite immune activation, has been proposed (McGonagle et al. 2001). Extra-articular involvement in the eye, heart and lung may be related to repeated cyclical biomechanical stressing at these sites. According to this model, two signals are required for the activation of the immune response in AS. The first signal comes from an antigen presenting cell displaying arthritrogenic peptides via HLA-B27 to an appropriate T cell. The second co-stimulatory signal comes from the local diseased microenvironment (McGonagle and Emery 2000). It has been proposed that the costimulatory signal in AS involves the activation of NF-κB, which is a pivotal transcription factor in chronic inflammatory disease (Barnes and Karin 1997). NF-κB can be activated by cytokines, such as TNF-α and IL-1β, protein kinase activators, oxidants, viruses, immune stimuli, antigens, bacterial heat shock proteins, bacterial CpG motifs and lipopolysaccharides (Barnes and Karin 1997; Yi et al. 1998; Kol et al. 1999). This activation in turn leads to the coordinated expression of many genes encoding cytokines, enzymes, adhesion molecules and chemokines, which can further amplify and perpetuate the inflammatory process (Barnes and Karin 1997).

Maksymowych has suggested that immunity to a component of fibrocartilage rather than enthesis is a more likely hypothesis in SpA pathology (Maksymowych 2000). The reason is that in AS pathologic changes frequently involve the fibrocartilaginous enthesis and cartilaginous joints, including the intervertebral discs, the manubriosternal joints and the symphysis pubis. The involvement of the hyaline cartilage of sacrum, vertebral endplate and hip has also been observed. An alternative explanation is that osteitis is primary and cartilage autoimmunity secondary (McGonagle and Emery 2000). Poole has suggested that commonly expressed autoantigens or common epitopes from structurally related cartilage proteins are responsible and cross-reactive immunity may explain the involvement of other tissues (the eye and arterial vessels) (Poole 1998). The G1 domain of aggrecan, the major cartilage proteoglycan has been suggested as an autoantigen and *in vitro* studies have demonstrated that a cellular immune response to G1 is present in most AS patients (Zou et al. 2003).

Recently, it has been proven that TNF-α antagonists such as infliximab or etanercept are efficacious in controlling signs and symptoms in patients with chronic AS (Brandt et al. 2000). Etanercept is associated with marked improvement of enthesitis and associated osteitis pathology as determined by MRI (Marzo-Ortega et al. 2001). Preliminary results suggest that a recombinant human IL-1 receptor antagonist (IL-1Ra) is also effective in AS (Haibel et al. 2003). This efficacy suggests a central role for proinflammatory cytokines such as TNF-α and IL-1 in the pathogenesis of the disease. Altogether, immune mediated mechanisms involving HLA-B27, inflammatory cellular infiltrates, cytokines and genetic and environmental factors are believed to have key roles in AS pathogenesis.

1.2 GENETICS OF AS

Twin and family studies indicate that the genetic contribution to the disease susceptibility accounts for >90% of the population variance (Brown et al. 1997b). Disease severity is also predominantly genetically determined (heritability 0.51-0.68) (Hamersma et al. 2001). The heritability as assessed by the excess sibling risk (λs) has been calculated ($\lambda s = Ks/K$) at 58 (see section 1.3.1) (Carter et al. 2000). The relative contribution of MHC and non-MHC genes in AS are discussed in sections 1.2.2 and 1.2.3, respectively.

1.2.1 HLA-B27

The discovery of the association between *HLA-B27* and AS dates from 1973 (Brewerton et al. 1973b; Schlosstein et al. 1973), it remains one of the strongest associations between an HLA antigen and a disease. Substantial evidence from *HLA-B27* transgenic animal and population studies strongly favours a direct role for *HLA-B27* in genetic susceptibility to AS. However, it remains an unproven hypothesis that *HLA-B27* itself is involved rather than a nearby gene.

HLA-B27 transgenic animal studies. *HLA-B27* transgenic rats with the appropriate genetic background develop a multisystem inflammatory disorder with many similarities to human SpA including gut, joint and skin lesions (Hammer et al. 1990). The disease is dependent on the presence of bacterial flora suggesting a critical involvement of microbes in the disease pathogenesis (Taurog et al. 1994). However, this model has significant differences to the human disease. Susceptibility to disease is related to the gene copy number and level of expression of HLA-B27, with disease developing only in those lines having supraphysiological levels of transgene expression (Taurog et al. 1993). In contrast to humans who develop axial disease, the rats develop a more prominent peripheral disease. *HLA-B27* transgenic rats develop orchitis and severe colitis, whereas human AS has no association with orchitis and severe colitis is uncommon (Hammer et al. 1990). Mice transgenic for *HLA-B27*, with or without human β2-microglobulin, have been described (Krimpenfort et al. 1987; Taurog et al. 1988; Khare et al. 1995). Unlike humans, these mice develop predominantly peripheral arthritis and enthesopathy.

Population studies. *HLA-B27* is strongly associated with AS world-wide, although the strength of the association varies between various ethnic groups (Calin and Taurog 1998). In Caucasians more than 90% of the AS patients have *HLA-B27* compared with only 4-14% of unaffected individuals. *HLA-B27* is virtually absent in most of the sub-Saharan Africa where AS is rare. In some of these areas *HLA-B27* negative AS is increased, the disease shows clinical features that are similar to those observed in Caucasian *HLA-B27* negative patients with AS. In these patients the disease onset is later compared with *HLA-B27* positive patients, the patients rarely get AAU and familial occurrence of AS is rarely observed (Mijiyawa et al. 2000). The epidemic of HIV infection in sub-Saharan Africa has been associated with a dramatic increase in the prevalence of SpA other than AS despite the low prevalence of *HLA-B27* (Mijiyawa et al. 2000). AS is virtually absent in West Africa where *HLA-B27* is more common in the general population (Brown et al. 1997a). The risk conferred by *HLA-B27* for AS is lower for Africans than that for black Americans (Mijiyawa et al. 2000).

Predictive value of HLA-B27 testing in the clinical practice. The penetrance of *HLA-B27* in AS is low, approximately 1-2% of *HLA-B27* positive adults in the general population are likely to develop AS (van der Linden et al. 1984b). Thus, when used in general population screening *HLA-B27* testing has a low positive predictive value – that is, the likelihood of disease is low when the test is positive. Bayesian modelling has demonstrated that the diagnostic value of *HLA-B27* testing in AS depends on the pre-test likelihood of the disease based on a clinical estimate. *HLA-B27* testing is useful if the pre-test probability is intermediate. For example, if the pre-test probability of AS in a Caucasian patient was 50%, a positive *HLA-B27* result would increase the likelihood that the patient has the disease to 92% and a negative result would decrease the likelihood that the patient has the disease to 8% (Khan and Khan 1982). These data are in accordance with a German study demonstrating that *HLA-B27*-positive blood donors with inflammatory back pain were 10 times more likely to have sacroiliitis and SpA than *HLA-B27*-negative donors (Braun et al. 1998). However, the clinical usefulness of *HLA-B27* typing as an aid to diagnosis of AS is much less in the black population due to weaker association between *HLA-B27* and disease.

Structure of HLA-B27. HLA-B27, in common with other MHC class I molecules, is a transmembrane glycoprotein expressed on the surface of most nucleated cells. The amino acid sequence of *HLA-B27* has been known since 1985 (Ezquerra et al. 1985). The HLA-B27 molecule has been crystallised and its structure determined (Figure 1.2) (Madden et al. 1991). HLA-B27 is associated with β2-microglobulin through its α3 domain and crystallographic studies indicate that the α1 and α2 domains constitute the peptide-binding site. The most distinctive structural feature is of HLA-B27 is its deep B pocket. Dissection of the peptide-binding grooves of the HLA-B27 subtypes shows that

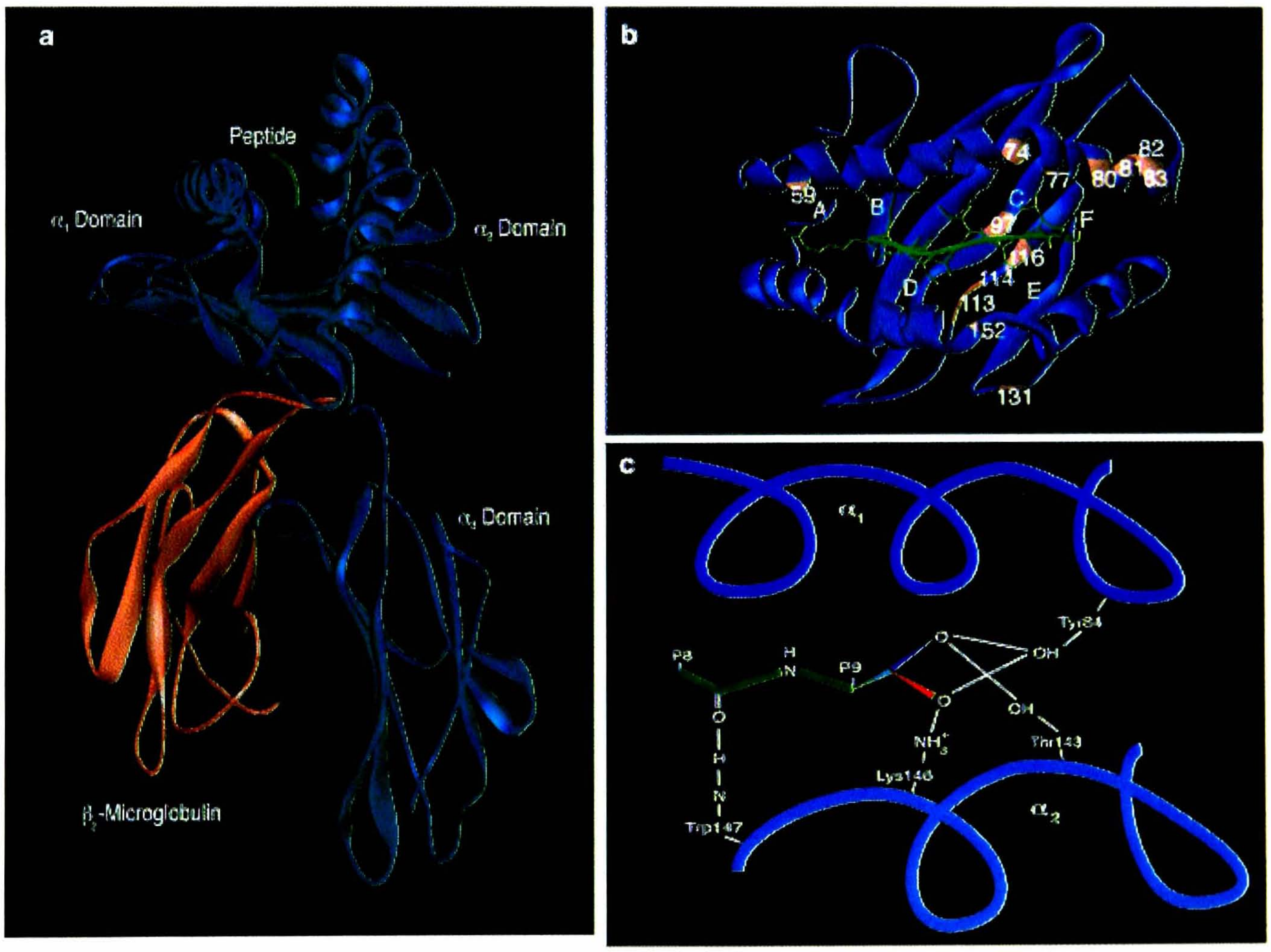
Figure 1.2 Schematic diagram of the three-dimensional structure of HLA-B27 molecule.

a) HLA-B27 heterodimer comprises of a heavy chain membrane glycoprotein of ~44 kDa noncovalently bound to a light chain of 12 kDa, termed β2-microglobulin.

b) Three-dimensional structure of the α1 and α2 domains of the B*2705. Numbers indicate polymorphic residues among HLA-B27 subtypes. Letters indicate the pockets of the peptide-binding groove. The peptide is shown in green and it interacts with the groove through hydrogen bonds.

c) Pocket F of HLA-B27. The residues of HLA-B27 that interact with position 9 of the binding peptide are shown.

Figure obtained from (Lopez-Larrea et al. 1998) with permission from Dr Carlos López-Larrea.



B pocket is conserved among all the common HLA-B27 subtypes (Buxton et al. 1992). HLA-B27 is highly selective for peptides of 9 amino acids with an arginine at the second position of the B pocket (Madden et al. 1991; Guo et al. 1993).

HLA-B27 subtype polymorphisms and association with disease. HLA-B27 itself is a serologic specificity, which encompasses a family of 24 different subtypes, denoted *HLA-B*2701 to B*2725.* (*HLA-B*2722* subtype has been eliminated as it turned out to have the same sequence as *HLA-B*2706*). The distribution of the subtypes varies strongly among different ethnic groups. *B*2705* is present in almost all populations of the world, but is most common in Circumpolar and Subartic regions from Eurasia and North America. *B*2702* has been described in Caucasians. *B*2704* and *B*2706* have been found mainly in Asians. *B*2703* is most common in West African populations. The structural pattern and ethnic distribution of the subtypes suggests that *B*2705* is the ancestral allele. Occurrence of at least one case of AS or related SpA has been reported in subjects possessing any of the first 10 subtypes, *B*2701* to *B*2710* (Gonzalez-Roces et al. 1997; Armas et al. 1999; Olivieri et al. 2002) as well as *B*2714* (Shankarkumar et al. 2002), *B*2715* (Garcia-Fernandez et al. 2001) and *B*2719* (Tamouza et al. 2001). However, only the relatively common subtypes, *B*2705, B*2702, B*2704* and *B*2707*, are proven to be disease associated by epidemiological case-control studies. Two subtypes, *B*2706* in Southeast Asia and *B*2709* in Sardinia lack a strong association with AS (Lopez-Larrea et al. 1995; Nasution et al. 1997). The differential disease associations between patients and healthy controls may provide clues to the effects of the sequence variations on the peptide binding specificity of the molecule.

The pathogenic link between *HLA-B27* and AS remains enigmatic. Different hypotheses have been formulated in the attempt to explain the association. A detailed evaluation of these theories is outside the scope of this thesis. The molecular mimicry theory was proposed in 1976, when it was shown that several gram-negative micro-organisms, such as *Klebsiella pneumoniae*, carry antigens that cross-react with HLA-B27 (Ebringer and Wilson 2000). This theory postulates that an autoimmune response initially mounted against an antigenic component of a micro-organism is directed against HLA-B27 itself. However, at least one controlled study failed to find a convincing association between microbial antigens and HLA-B27 (Lahesmaa et al. 1992), and human studies remain inconclusive (Ringrose 1999). Additionally, this theory cannot explain the differential *HLA-B27* subtype associations.

A related hypothesis, the arthritogenic peptide model predicts that HLA-B27, as a result of its particular peptide binding specificity, presents self-peptides mimicking pathogen-derived epitopes that then become the target of autoreactive cytotoxic T lymphocytes (CTL) (Benjamin and Parham 1990). In support of this hypothesis, it has been demonstrated that HLA-B27 is capable of presenting potentially arthritogenic peptides to CTL. Hermann et al. isolated HLA-B27-restricted autoreactive CD8+ T cells from synovial fluids of patients with reactive arthritis and AS (Hermann et al. 1993). It has also been suggested that amino acid differences in *B*2706* and *B*2709*, which limit their peptide repertoire compared with *B*2705*, also prevent the presentation of putative arthritogenic peptides. Recently HLA-B27 reactive CD4+ T cells have been identified in the *HLA-B27*-positive AS patients (Boyle and Hill Gaston 2003). These cells may play a role in the pathogenesis of SpA.

New ideas have focussed on abnormal characteristics of HLA-B27 resulting from disulfide bond formation, aberrant folding or both. It has been recently demonstrated that HLA-B27 has an unusual ability to form heavy chain homodimers *in vitro* and in certain cell lines *in vivo* (Allen et al. 1999). Dimerization *in vitro* is dependent upon the formation of a disulphide bond through unpaired cysteines at position 67 in the α1 helix of the HLA-B27 heavy chain. These homodimers lack β2-microglobulin. HLA-B27 heavy chain homodimers and receptors for HLA-B27 homodimers are expressed on populations of peripheral blood and synovial monocytes and B and T lymphocytes from patients with SpA (Kollnberger et al. 2002). Control subjects also express receptors for HLA-B27 heavy chain homodimers. One possible mechanism by which expression of HLA-B27 heavy chain homodimers could lead to joint inflammation is presentation of peptides to either CD4+ or CD8+ T cells, or more likely interaction with natural killer (NK) family receptors (Kollnberger et al. 2002).

An alternative hypothesis relates to the observation that B*2705 heavy chain folded much more slowly than a mutant in which six B pocket residues had been replaced by those in HLA-A*0201, resulting in misfolding of a portion of HLA-B27 pool (Mear et al. 1999). These results indicate that in addition to its role in peptide selection, the B pocket is also involved in misfolding a portion of HLA-B27 heavy chains in the endoplasmic reticulum (ER), resulting in their degradation in the cytosol. On the basis of these observations, it was suggested that an accumulation of misfolded B*2705 in the ER might elicit an ER stress response. This could lead to the activation of NF-κB, which in appropriate cell types might increase the production of proinflammatory cytokines. Colbert proposed that even if the degree of misfolding is insufficient to stimulate cytokine production, it could lower the threshold of activation by other stimuli, such as bacterial infection (Colbert 2000).

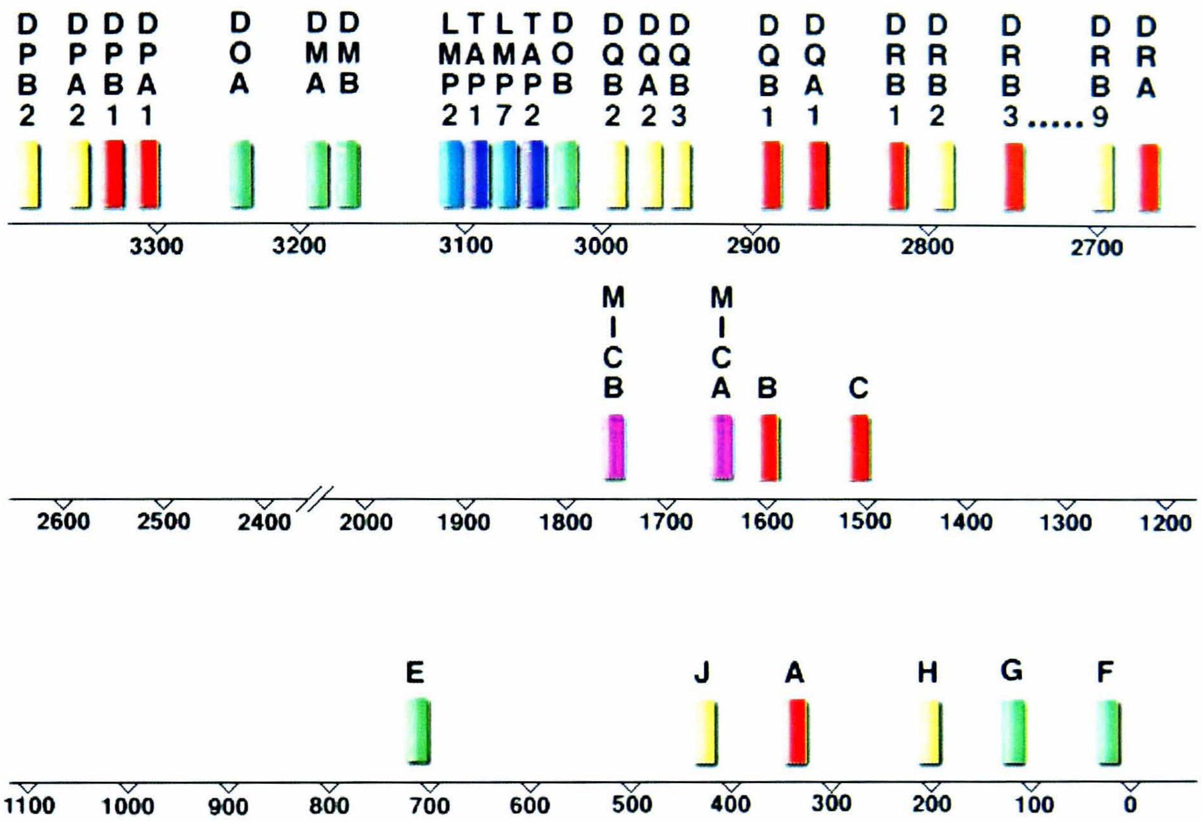
1.2.2 MHC genes and AS

Genes of the human MHC, in particular *HLA-B27*, are major genetic factors in determining the familial clustering of AS, as evidenced by linkage and association studies (Brewerton et al. 1973b; Schlosstein et al. 1973; Rubin et al. 1994; Laval et al. 2001). A recent whole genome screen yielded a maximum LOD score of 15.6 for the MHC region indicating that it is the major susceptibility locus in AS (Laval et al. 2001). *HLA-B27*, although almost essential for the inheritance of AS, is estimated to contribute only 16-50% to the overall genetic component (Brown et al. 1997b; Carter et al. 2000). *HLA-B27* positive relatives of AS patients have a recurrence risk of the disease that is 5.6-16 times higher than that for *HLA-B27* positive individuals in the general population, implying the presence of non-*HLA-B27* shared familial risk factors (Calin et al. 1983; van der Linden et al. 1984b). Several haplotypic and association studies suggest that more than one gene within the MHC may influence the disease susceptibility (Brown et al. 1998a; Hohler et al. 1998).

Figure 1.3 Graphical representation of the MHC region.

Picture obtained from the Anthony Nolan database (www.anthonynolan.org.uk/HIG/index.html).

The numbers underneath refer to the kilobases.



Other HLA class I antigen associations have been studied extensively but only *HLA-B60* association has been confirmed by several independent groups. *HLA-B60* is associated with 2-4 times increased risk of AS in Caucasian populations (Robinson et al. 1989; Brown et al. 1996b). An association between AS and *HLA-B*1403* has been noted in a West African population (Lopez-Larrea et al. 2002). A recent Mexican study reported a moderate increase of *HLA-B15* among patients with SpA (Vargas-Alarcon et al. 2002b). *HLA-B39* has been reported to be associated with AS in an *HLA-B27* negative Japanese population (Yamaguchi et al. 1995), although not in other populations (Brown et al. 1996a). A small Spanish study demonstrated a significant increase in the *HLA-A9* allele frequency in AS patients compared to an *HLA-B27* positive control group (de Juan et al. 1999).

The *MICA* gene is located 47 kb centromeric from the *HLA-B* gene and is expressed in gastrointestinal epithelium, fibroblasts and endothelial cells (Groh et al. 1996; Zwirner et al. 1999). It has been suggested that MICA is recognised by diverse $V_{\delta}1$ $\gamma\delta$ T cells in the epithelium of intestinal tract (Groh et al. 1998). A Sardinian study demonstrated an association between *MICA-A4* allele and *HLA-B27* negative AS patients (Ricci-Vitiani et al. 2000), but the sample size for this analysis was small. However, several studies have observed no association between polymorphisms of the *MICA* gene and AS independent of *HLA-B27* (Tsuchiya et al. 1998; Yabuki et al. 1999; Martinez-Borra et al. 2000), suggesting that at least in *HLA-B27* positive individuals *MICA* alleles are not likely to play a role.

Several previously published studies have demonstrated weak associations between *HLA-DRB1* alleles and both disease susceptibility and clinical manifestations of the disease. These studies are discussed in chapter 5. Other MHC class II genes encoding products involved in class I antigen processing pathway have been investigated in AS. These include the transporter associated with antigen processing *(TAP)* genes, *TAP1* and *TAP2*, and the large-molecular-weight proteasome *(LMP)* genes, *LMP2* and *LMP7*. Iritis and extraspinal disease in AS have been reported to be associated with polymorphisms in *LMP2* gene (Maksymowych et al. 1997c; Maksymowych et al. 2000), but other studies have not confirmed this (Burney et al. 1994). A Spanish study noted an association between an *LMP7* gene polymorphism and susceptibility to AS (Fraile et al. 1998b). No significant association has been reported between *TAP1* and *TAP2* gene variants and SpA (Burney et al. 1994; Westman et al. 1995; Fraile et al. 2000).

HLA class III. *TNF-α* gene is located 250 kb centromeric of *HLA-B*. *TNF-α* mediates inflammation and has immunoregulatory activities. The production of TNF-α is highly heritable (Westendorp et al. 1997), and several studies have shown that TNF-α levels are raised at baseline in patients with SpA. Gratacos et al. analysed the cytokine profiles in the serum from 69 patients with AS and 36 controls with non-inflammatory back pain using an enzyme linked immunosorbent assay (ELISA) (Gratacos et al. 1994). TNF-α levels were raised in patients with AS compared with patients with non-inflammatory back pain ($p=0.016$). Sonel et al. compared cytokine levels between 22 healthy controls and 42 patients with SpA having either active or inactive disease (Sonel et al. 2002). Both SpA groups with active and inactive disease had higher serum TNF-α levels than controls ($p<0.05$). High amounts of TNF-α mRNA were observed in biopsy specimens of sacroiliac joints in patients with AS (Braun et al. 1995), demonstrating that TNF-α acts at the primary site of inflammation. AS patients from Germany and western Scotland have a decreased frequency of the *TNF-α* -308.2 allele (Hohler et al. 1998; McGarry et al. 1999), although this association has not been demonstrated in British or Dutch patients (Kaijzel et al. 1999; Milicic et al. 2000). The discrepant association between different ethnic populations possibly suggests that this *TNF-α* polymorphism may merely be in LD with another disease-causing gene rather than being the true disease causing variant itself, as suggested by the different *HLA-DRB1-B27* LD patterns in British Caucasians and Germans (Milicic et al. 2000).

Heat shock protein 70 (HSP70) loci are located between the complement and tumour necrosis factor genes and include three immediately adjacent genes, *HSP70-1*, *HSP70-2* and *HSP70-hom*. HSP70 proteins have been linked with a protective role during and after cellular stress (Sargent et al. 1989). Finnish and Spanish studies demonstrated no independent association between *HSP70* gene polymorphisms and AS (Westman et al. 1994b; Fraile et al. 1998a). A recent Mexican study noted a significant association between *HSP70-hom* A allele and A/A genotype between *HLA-B27* negative SpA patients and healthy controls (Vargas-Alarcon et al. 2002a), but the control genotypes in this study were not in Hardy-Weinberg equilibrium (HWE), suggesting either a sampling bias or a genotyping error.

1.2.3 Non-MHC genes and AS

The genetic contribution of genes lying outside the MHC region has been investigated by a systematic genome-wide approach and by functional and positional candidate gene analysis. A large non-MHC component of the genetic risk for the disease is suggested by the significant difference in concordance rates for monozygotic (MZ) twins (63%) and *HLA-B27*-positive dizygotic (DZ) twins (23%) (Brown et al. 1997b). Whole-genome screening of AS showed suggestive or stronger linkage with the disease on chromosomes 1p, 2q, 9q, 10q, 16q and 19q (Laval et al. 2001). Recurrence risk modelling suggests that the best fitting model of genetic susceptibility to AS is a five-locus model with multiplicative interaction between loci (Brown et al. 2000b). The susceptibility of AS is likely to arise from the combined impact of multiple contributing genes, each potentially interacting with environmental and stochastic factors.

Recent studies have added to the evidence that cytokines play an important role in AS pathogenesis. Several candidate gene studies have investigated the effect of cytokine gene polymorphisms in AS. Recently intracellular cytokine staining demonstrated that patients with AS and healthy *HLA-B27* positive controls produced significantly less

TNF-α and interferon γ (IFNγ) than *HLA-B27* negative controls (Rudwaleit et al. 2001). In contrast, production of IL-10 by CD8+ T cells was higher in patients with AS than in *HLA-B27* positive and negative controls. IL-10 is a potent inhibitor of many proinflammatory cytokines, it attenuates the elimination of a variety of pathogens thus influencing the outcome of disease (Redpath et al. 2001). It has been estimated that 75% of the variation in IL-10 production is genetically determined (Westendorp et al. 1997). Haplotypes defined by the microsatellites IL10.G12 and G10 in the *IL-10* promoter associated with a protective effect against the development of reactive arthritis in Finnish patients (Kaluza et al. 2001), suggesting that *IL-10* promoter region variants may play a role in SpA. However, no association between *IL-10* promoter variants and susceptibility to disease was noted in Finnish families with AS (Goedecke et al. 2003). *TGFB1* is another immunosuppressive cytokine, and *TGFB1* polymorphisms in AS are discussed in chapter 3.

The *IL-1* gene cluster is located on chromosome 2q13-14. It encodes for three proteins, IL-1α, IL-1β and interleukin-1 receptor antagonist IL-1Ra. The first two are strong inducers of inflammation, while IL-1Ra binds to the IL-1 receptors with high affinity without activating the cell and is thus an effective antagonist (Hurme and Santtila 1998). This cluster is an interesting positional and functional candidate region for AS, as this locus is approximately 0.3 cM centromeric to the marker D2S160 which is strongly linked to AS (p=0.007) (Laval et al. 2001). Polymorphisms in these cytokine genes have been implicated in several other inflammatory and autoimmune diseases (Blakemore et al. 1995; Heresbach et al. 1997; Schrijver et al. 1999). Allele 2 in the *IL-1Ra* has been associated with susceptibility to AS by a Scottish and a Dutch group (McGarry et al. 2001b; van der Paardt et al. 2002a), but this was not confirmed by a small French study (Djouadi et al. 2001). Allele 2 of the *IL-1Ra* is associated with increased production of IL-1Ra protein (Danis et al. 1995; Hurme et al. 1998), suggesting a functional role of this variant. Recently, an association between AS and single nucleotide polymorphisms (SNPs) in the 3' region of the *IL-1Ra* gene was reported in a large case-control and family-based study (Maksymowych et al. 2003). Given these three independent reports of positive association between variants in this gene and AS, it is very likely that either *IL-Ra* gene itself or another tightly linked gene is a true susceptibility gene in AS. Further studies are required to pinpoint the causative variant/s at this locus.

IL-6 is a proinflammatory cytokine involved in local inflammatory reactions by amplifying leukocyte recruitment (Romano et al. 1997). *IL-6* gene is located on chromosome 7p21. IL-6 levels are increased in patients with AS and a close correlation between serum IL-6 and disease activity and severity has been noted (Gratacos et al. 1994). A relatively small Spanish study noted no association between *IL-6* -174 polymorphism and susceptibility to AS (Collado-Escobar et al. 2000).

The *NOD2/CARD15* gene is located on 16q12. Recently an association between the variants of the *NOD2* gene and Crohn's disease has been discovered (Hugot et al. 2001; Ogura et al. 2001). NOD2 is involved in the inflammatory response to bacteria via the activation of NF-κB pathway, following its interaction with bacterial components. Several studies have concluded that polymorphisms in the *NOD2* gene are not involved in susceptibility to AS (Crane et al. 2002; D'Amato 2002; Miceli-Richard et al. 2002; Ferreiros-Vidal et al. 2003). A British study reported an association between the carriage of *NOD2* Pro268Ser polymorphism and greater disease activity as measured by the BASDAI (p=0.002) (Crane et al. 2002), indicating that this locus may act as a severity factor.

The human analogue of the mouse progressive ankylosis (*ank*) gene, *ANKH* is a candidate gene for AS based on the mouse model. Mice homozygous for the Glu440X mutation in the *ank* gene develop a generalised progressive form of arthritis accompanied by mineral deposition, aberrant new bone formation resulting in ankylosis and eventual joint fusion (Ho et al. 2000). The *ANKH* gene has been mapped to 5p14.1-p15.2. Mutations in *ANKH* have been implicated in autosomal dominant craniometaphyseal dysplasia and in familial calcium pyrophosphate deposition disease (Nurnberg et al. 2001; Reichenberger et al. 2001; Pendleton et al. 2002; Williams et al. 2002). A recent study reported significant associations between the promoter polymorphisms of the *ANKH* gene and AS (Tsui et al. 2003), but this was not confirmed by a large British study (Timms et al. 2003).

1.3 GENETICS OF COMPLEX DISEASES

1.3.1 Estimating genetic risk

Risk ratio of the disease. Calculating risk ratios between affected individuals and their relatives (e.g. siblings) is a powerful way of estimating the strength of an underlying genetic effect of the disease. Sibling relative risk ratio λs is commonly used and is calculated as follows:

$\lambda s = Ks/K$

in which Ks is the probability that a sibling of an affected individual is also affected and K is the population frequency of a disease. λs defines the ratio of increase in risk of disease given the sibling has the disease. λs may substantially overestimate the genetic contribution, because siblings are likely to share the same environment. Because λ is a ratio, the prevalence of the disease among the relatives and the prevalence of the disease in the general population affect its size. The risk ratio λ decreases with the degree of relationship between proband and relatives, and the rate of decrease depends on the mode of inheritance underlying the trait (Risch 1990a).

Heritability. Heritability refers to the proportion of phenotypic variation that can be ascribed to genotypic variation. Heritability can be determined using family and twin studies. Since MZ twins share all their genes whereas DZ twins are assumed to share on average 50% of their genome, the ratio of their concordances for a condition gives a rough indication of the relative importance of genotype in its causation. Both groups are assumed to share environmental factors to a similar extent. A disadvantage is that twins are difficult to recruit especially for relatively rare diseases. Also, twin studies are a statistically weak method for investigating genetic dominance.

Risk ratio of the disease locus. Once genetic susceptibility loci are identified, locus specific λs can be determined by calculating the ratio of expected proportion of affected sibling pairs sharing zero alleles identical by descent (0.25) by the actual observed fraction. Locus specific λs facilitates estimation of the number of major genes remaining to be found, although the interaction between these genes may be multiplicative. Another measure of the effect of a locus is the parameter γ, the genotype relative risk (GRR) associated with heterozygosity or homozygosity for a disease susceptibility allele (Risch 2000).

1.3.2 Linkage analysis

Definition. Linkage analysis relies on the fact that loci which are physically adjacent to each other in the same chromosome are likely to be inherited together, the closer the loci are the more rarely they are separated by genetic recombination. It aims to identify regions of the genome that co-segregate with the disease in many independent families or over many generations in an extended pedigree. Genes can usually be localised to a large physical interval by this approach as the size of the co-segregating segment of DNA is only delimited by the observation of a crossover between a marker and the disease locus itself. The probability that recombination will occur between two loci is, in part, a function of the physical distance separating them. However, this relationship is not linear for several reasons: (1) recombination results from an odd number of crossovers, which are not directly observed; (2) the density of occurrence of crossover appears to be variable along the chromosomes; (3) multiple crossovers do not appear to occur independently (Lalouel and White 1997). A crossover event tends to inhibit the occurrence of another in its vicinity, a phenomenon called positive interference.

Map functions. A mathematical relationship that converts recombination fraction to genetic map distance is called a map function. Haldane's function assumes that crossovers occur at random along a chromosome and they do not influence on one another. Haldane's function is expressed as:

$m = -0.5 \ln (1-2\theta)$,

in which m is the map distance and θ is the recombination fraction (Haldane 1919). Kosambi's mapping function takes into account interference and is described by:

$m = 0.25 \ln [(1+2\theta)/ (1-2\theta)]$ (Kosambi 1944).

For small values of θ (<0.1) the Haldane and Kosambi mapping functions provide a similar estimate of genetic distance, which is equivalent to the recombination fraction expressed as a percentage. For greater recombination values, interference becomes more relevant and the estimates of genetic distance provided by these functions can differ.

Genetic distance vs. physical distance. Genetic distance between loci is measured in centiMorgans (cM) and 1 cM represents the distance associated with a 1% recombination fraction (θ=0.01) between two loci. The relationship between genetic and physical distance is not constant. On average, a genetic distance of 1 cM equates to a physical distance of 1 Mb of DNA sequence, but this varies markedly between different genetic regions. The genetic distance/physical distance differs considerably between genders, as recombination frequencies are higher in females than in males

(Broman et al. 1998). To overcome these differences, genetic maps typically report distances for each sex and a sex-averaged distance that integrates male and female recombination frequencies. In females, the crossover rates are much higher around the centromeres, whereas those in males tend to be higher towards the telomeres (Broman et al. 1998).

Linkage maps. Genetic linkage maps measure the underlying genetic recombination that occurs in a particular chromosomal region. They can also serve as frameworks for building physical maps. Construction of precise genetic linkage maps is important as the precise estimates of genetic distance improve the statistical power of linkage analysis. With actual linkage, any map misspecification causes negative bias in multipoint (analysis of several markers at once) LOD scores, resulting in loss of power to detect linkage. In the absence of linkage, map misspecification can cause positive or negative bias: using too large a distance gives a positive bias; using too small a distance gives a negative bias; falsely assuming a 1:1 female:male ratio always causes positive bias (Daw et al. 2000). A range of genome-wide human genetic maps has been published to date. The most recent genetic map based on microsatellite markers has been developed by the deCode genetics (Kong et al. 2002), and it provides higher resolution compared to the Marshfield map (Broman et al. 1998). Recently, a genome-wide SNP linkage map has been published based on 2679 SNP markers (Matise et al. 2003). The disadvantage of using SNP based linkage maps is the lower information content of a single SNP compared to a polymorphic microsatellite, but this can be overcome by increasing the marker density 3-8-fold (Kruglyak 1997), or using multiple clusters of two or three SNPs in linkage analysis at a similar density to microsatellites (Goddard and Wijsman 2002).

LOD scores. Linkage is statistically calculated by the LOD score method (Morton 1955). It gives the statistical likelihood that the loci are linked with recombination fraction θ compared with the likelihood that they are inherited independently, when recombination is observed in 50% of the meioses, corresponding to θ=0.5. The ratio of these likelihoods indicates the odds for or against linkage and is presented as a LOD score estimate as follows: $LOD(\theta) = \log_{10} (L(\theta)/ L(1/2))$.

For a Mendelian character, a LOD score greater than 3 has been deemed to indicate significant linkage. However, it is of note that even a LOD score of 3 will prove to be spurious in approximately 1/20 instances (p=0.05) (Nyholt 2000). A LOD score of -2 or less is considered to exclude linkage. Lander and Kruglyak proposed the following criteria for interpretation of linkage findings of complex traits (Lander and Kruglyak 1995):

1. *Suggestive linkage* – statistical evidence expected to occur one time at random in a genome scan
2. *Significant linkage* – statistical evidence expected to occur 0.05 times in a genome scan (ie. with the probability 5%)
3. *Highly significant linkage* – statistical evidence expected to occur 0.001 times in a genome scan
4. *Confirmed linkage* – significant linkage from one or a combination of initial studies that has subsequently been confirmed in a further sample, preferably by an independent group of investigators. A nominal p-value of 0.01 should be required for confirmation. In case of sib pair studies, the first three categories would correspond to pointwise significance levels of 7×10^{-4}, 2×10^{-5} and 3×10^{-7}, and LOD scores of 2.2,

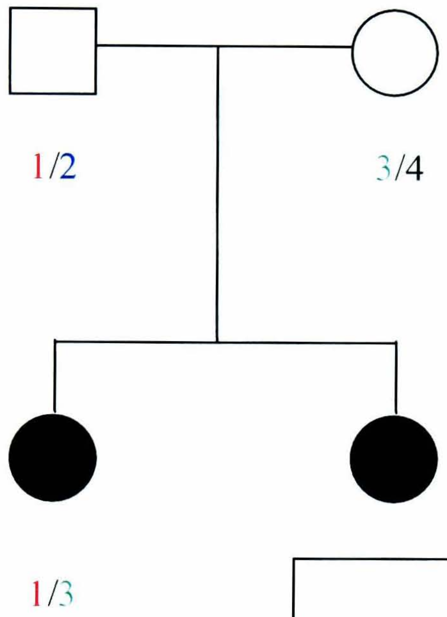
3.6 and 5.4. Nevertheless, LOD scores may vary greatly in their statistical properties and, consequently, in their level of significance depending on the statistical test employed (Nyholt 2000).

Parametric linkage analysis. Parametric linkage analysis has proved a powerful way to map the genes underlying monogenic disorders. The basis of the parametric linkage approach is to search for recombinants and non-recombinants between the phenotype and the marker being tested. If a marker is unlinked to a disease, each child will have an equal chance of receiving each of the parental alleles. If the marker is linked to the disease, offspring will be more likely to inherit one allele over the other, depending on whether or not they are affected. Parametric linkage analysis requires the specification of an inheritance model, including gene frequencies and penetrance. The power to detect linkage by parametric LOD score analysis is sensitive to misspecification of the genetic model: the false negative (but not false positive) rate increases with model misspecification. If the mode of inheritance is known, traditional parametric methods are more powerful than non-parametric linkage (NPL) methods.

NPL analysis. NPL analysis does not require the specification of values for the parameters defining the transmission model and allele frequencies; thus it is suitable to disease gene mapping of complex disease. The commonly employed study design of NPL mapping requires affected sibling pairs and their parents (figure 1.4). This affected-sib-pair test compares the observed number of affected sibling pairs sharing zero, one or two alleles identical by descent (ibd) with that which is expected with no linkage (proportioned 25%, 50%, 25%) (Nyholt 2000). If the marker is linked to the disease locus, the affected sibling pairs will share the disease allele ibd more often than they would a marker unlinked to the disease locus.

Figure 1.4 Affected-sib-pair test.

The probabilities of two siblings sharing 0, 1 or 2 alleles ibd are 0.25, 0.5 and 0.25, respectively. Ibd sharing may be equivocal, if both parents are not heterozygous with different genotypes.



	Alleles shared	Probability
2/4	0	1/4
2/3 or 1/4	1	1/2
1/3	2	1/4

Power of linkage analysis. The critical parameters in determining power to detect linkage by using affected pairs of relatives are the magnitude of underlying genetic effect, and recombination fraction between the trait and the marker loci (Risch 1990b). Reduced marker polymorphism results in decreased power (Risch 1990c). For a partially penetrant recessive trait, affected sib pairs are the most informative type of relative pair. For a dominant trait, other relationships may be considerably more informative, depending on allele frequency of the trait and marker polymorphism. Risch and Merikangas demonstrated that linkage analysis is likely to be successful only for loci with γ (GRR) values in the range of four or larger, provided the allele frequency is between 5 and 75%, but not for loci with γ values of 2 or less (Risch and Merikangas 1996). Linkage analysis can fail even in very large studies, if misdiagnoses, heterogeneity, complex inheritance or frequent phenocopies are abundant (Botstein and Risch 2003). Genome-wide linkage scans designed to localise disease genes for complex diseases have yielded few significant findings. Failure to reproduce linkage results in subsequent studies is frequent.

1.3.3 Association analysis

Definition. Association analysis looks for a statistical correlation or dependence between genetic variants and a disease on a population scale. Distantly related individuals are separated by numerous random recombination events. Association studies draw from historic recombination so disease-associated regions are very small in random mating populations, usually encompassing only one gene or gene fragment (Cardon and Bell 2001). As the disease mutation is transmitted through generations, recombination will separate it from the specific alleles of its original haplotype. Particular DNA variants can remain together on ancestral haplotypes for many generations; this creates the non-random association of alleles. Association studies can be especially useful for narrowing a complex trait candidate interval identified by linkage analysis and are commonly used for fine mapping of the disease causing gene. Statistical evidence for an association between an allele and a phenotype arises in one of three situations. First, causal variants may be associated with the disease phenotype. Second, linked neutral variants may be associated with the disease phenotype through

LD between the marker allele and the causal variant. Third, the association can be simply attributable to artefacts such as population stratification.

Power of association studies. Association studies are likely to be more effective than linkage studies for studying complex traits because they can have greater statistical power to detect several genes of small effect. Risch and Merikangas showed that even if one tests 1,000,000 polymorphic alleles genome-wide and allows for a conservative significance level of 5×10^{-8}, gene effects with γ values of as low as 1.5 could be detected in realistically sized (<1000 families) association studies (Risch and Merikangas 1996). They assumed that the polymorphisms tested are disease predisposing or in complete LD with the disease causing variant.

In case-control analysis the critical factor influencing test power is simply the difference in allele frequency between the cases and controls. This depends on LD between the marker alleles and the disease locus alleles and the strength of association between the disease mutation and the disease phenotype. The precise magnitude of the detectable effect at the particular locus will depend importantly on the sample size. With increasing sample size, alleles of increasingly minor effects can be identified (Ardlie et al. 2002a). The statistical power is also influenced by the frequency of the alleles of interest. Association studies have more power when the predisposing variant is frequent (Botstein and Risch 2003). This observation has fuelled much speculation on the likely frequency of the allelic variants for common disease (Pritchard 2001; Pritchard and Cox 2002; Lohmueller et al. 2003). SNPs with very low allele frequencies would need to have very large relative risks associated with them to be detected in an association study.

The power to detect a disease susceptibility locus decrease with diminishing LD between a tested marker and the disease locus (Muller-Myhsok and Abel 1997). When the marker allele is in LD with the disease locus, the power is determined by the extent of LD between the marker and the disease locus and the potential discrepancy between the frequency of the disease allele and the associated haplotype.

The LD measure r^2 represents the statistical correlation between two sites, and it is calculated as follows: $r^2 = (D_{AB})^2 / f_A(1-f_A)f_B(1-f_B)$ (Hill and Robertson 1968), in which $D_{AB} = f_{AB} - f_A \times f_B$

and f_{AB} represents the observed frequency of gamete AB, and f_A and f_B represent the frequencies of alleles A and B. r^2 is considered to be the most relevant measure for association mapping, because there is a simple inverse relationship between r^2 and the sample size required to detect association between susceptibility loci and SNPs. In order to have the same power to detect the association between the disease locus and the marker locus, the sample size must be increased by roughly $1/r^2$ times when compared with the sample size for detecting association with the susceptibility locus itself (Kruglyak 1999; Pritchard and Przeworski 2001). As a rough rule of thumb, r^2-values of above 1/3 might indicate sufficiently strong LD to be useful for mapping (Ardlie et al. 2002a).

The power of association studies is high when the allele frequency of the marker allele and the disease allele are equal or similar. In case of discrepant allele frequencies, the sample size must be increased by a factor of r, which is the ratio of the associated haplotype frequency to the disease allele frequency (Botstein and Risch 2003). For example, if the disease allele had a frequency of 5% and the associated haplotype had a frequency of 25%, a sample size 5 times larger would be required for the haplotype approach.

Case-control studies. The traditional case-control approach is the simplest study design for assessment of genotype-phenotype correlation. The case-control design employs simple statistics, and the recruitment of unrelated patients is easier than families especially in the case of late-onset diseases. Case-control studies are more powerful than family-based linkage analysis for detection of weak genetic effects and are more cost-efficient in terms of recruitment and genotyping effort (Risch and Teng 1998). However, when using this approach the association may be due to errors introduced by selection bias, recall bias, misclassification or confounding. Selection bias is the result of systematic differences between those who are selected for study and those who are not. Recall bias results from systematic differences in the accuracy or completeness of recall of past exposes or family history. Misclassification results from errors in the classification of individuals by phenotype, exposures or genotype. Confounding refers to the distortion of the measure of association because of the association of other non-intermediate factors with both the variable of interest and the outcome of interest (Tabor et al. 2002).

Family-based association studies. The most widely employed approach to improve the genetic matching of cases and controls includes use of relatives as controls. Several statistics have been formulated to analyse association within families. Falk and Rubinstein proposed a method called haplotype relative risk (HRR), which compares the genotype frequencies in the affected children with the non-transmitted parental genotypes frequencies (Falk and Rubinstein 1987). Terwilliger and Ott further developed this statistic to consider allele frequencies, rather than genotype frequencies, observed in the affected children versus the non-transmitted parental alleles (Terwilliger and Ott 1992). They referred to this statistic as haplotype-based relative risk (HHRR). Another statistic, the transmission disequilibrium test (TDT) is based on the fact that heterozygote parents have a 50% probability of transmitting an allele to their offspring in the absence of linkage and association (Spielman et al. 1993). In the presence of an association between a genetic marker and disease susceptibility locus, significant over-transmission of the associated allele can be noted. TDT can be extended to families with more than one affected child, because under the null hypothesis of no LD, every child in a family has an independent probability of 50% of inheriting either of the two alleles from a heterozygous parent. However, when families with more than one affected child are used, TDT is affected by the strength of linkage and the statistic is no longer a measure of association. Several modifications of the original TDT test have been developed, including sib-TDT test (Spielman and Ewens 1998) and methods suitable for situations where transmission is uncertain (Clayton 1999). Affected family based controls (AFBAC) test compares the frequency of alleles transmitted to affected children versus alleles never transmitted to an affected child (Thomson 1995). The relative power of these test statistics depends on the precise genetic model and population mating pattern. The advantages of family-based designs are the potential to conduct linkage analysis, assessment of parent-of-origin differences, and genotype-phase inference for haplotyping and assessment of genotyping errors. The disadvantage of a family-based approach is the relatively high cost of genotyping and low efficiency of recruitment.

Controlling for stratification with anonymous genetic markers. In case-control studies of association, population subdivision or recent admixture of populations can lead to spurious associations between a phenotype and candidate loci. Such associations arise because population subdivision permits marker allele frequencies to vary between segments of the population as the result of genetic drift or founder effects. Several methods have been designed to recognise the situation when structure may affect a case-control study and to correct for it when it is present. Pritchard and Rosenberg propose that anonymous genetic markers scattered throughout the genome indicate the amount of background diversity in cases and controls (Pritchard and Rosenberg 1999). They proposed using markers that do not correlate with each other and are independent of those affecting the disease of interest. As few as 30 SNPs are sufficient to detect population substructure. Once significant population stratification is detected, other statistical methods are required to control for its effect. Genomic control -method uses independent marker loci to adjust the distribution of the standard chi-squared statistic by a multiplicative factor that is proportional to the degree of stratification (Devlin et al. 2001). Another approach, called structure assessment, defines underlying subgroups in stratified samples, measures association in each stratum and combines the results of each subgroup (Pritchard et al. 2000; Pritchard and Donnelly 2001).

Forward genetics - Candidate gene approach. Candidate gene studies identify biologically meaningful genes for the disease of interest (Tabor et al. 2002). Generally candidate genes are prioritised on the basis of probable functional significance and appropriate expression profile. Polymorphic variants within and nearby the candidate genes are identified and genotyped in a set of cases and controls, and association between the disease phenotype and the markers are assessed. The candidate gene approach is commonly used. Disadvantages of this approach are the multitude of possible candidates genome-wide and the paucity of functional data available for most genes. Candidate gene numbers can be limited to those found in genomic regions identified by prior linkage studies, this is 'the positional candidate' approach.

Reverse genetics - Positional disease gene mapping using LD approach. The principle behind LD mapping is that variants that are not present in the initial screen can be investigated indirectly through LD with adjacent markers. LD mapping has been successful in identifying several genes for monogenic disorders inherited in Mendelian pattern (Hastbacka et al. 1992; Hastbacka et al. 1994) and recent success of identifying a Crohn's disease associated gene, *NOD2/CARD15*, demonstrates the utility of LD mapping in complex disease (Hugot et al. 2001; Ogura et al. 2001).

Population specific LD pattern and interethnic differences. The strength of the disease association may vary depending on gender, age and ethnicity (Farrer et al. 1997). Modifier genes and environmental triggers may contribute to the variations in phenotypic expression of the disease. A causal association between a candidate SNP and a disease should be reproducible in many ethnically diverse populations. Consistent replication in different populations is strong evidence of causality. However, the frequency of genetic disorders varies remarkably across populations, and not all SNP associations are universal. The absence of pan-ethnic replicability of association does not necessarily negate a causal relationship (Risch 2000). It may indicate that further studies in certain populations are required or for more detailed study of the function of the implicated gene. Patterns of LD may differ ethnically, helping to resolve causal from non-causal relationship. European populations typically show greater LD than

African populations (Reich et al. 2001) and it has been suggested that Caucasians would be more suitable for initial LD mapping of a disease gene. African populations exhibit much shorter range of LD and may be suitable for fine-mapping the disease causing variant.

Genetic mapping in isolated populations. Isolated populations established by a limited number of founders have proven extremely useful for mapping genes for rare monogenic disorders. Hastbacka et al. adapted Luria and Delbruck's methods of analysing bacterial cultures to the study of a Finnish founder population to estimate the recombination fraction between a diastrophic dysplasia locus and a marker (Hastbacka et al. 1992). This led to the identification of a defect in the causative gene encoding a sulphate transporter (Hastbacka et al. 1994). Following this success, it was thought that the same would hold for common diseases. Simulation studies have suggested that the value of isolates for common diseases may have been overrated (Kruglyak 1999). Empirical evidence suggests that several populations, such as the Finnish and Sardinians that were assumed to have high LD due to their founder history and relatively small size, do not appear to have greatly increased LD (Eaves et al. 2000; Taillon-Miller et al. 2000). Very small, isolated populations, exemplified by the Saami, have high levels of LD (Laan and Paabo 1997), but due to their small population size it is difficult to recruit a large enough patient cohort for genetic studies of a complex disease. In populations with relatively few founders, the genetic basis of disease susceptibility is likely to be less heterogenous than in large populations. Genetic isolates exhibit reduced number of polymorphic SNPs, for example 13% and 56% of SNPs are nonpolymorphic in Finns and Sardinians respectively (Taillon-Miller et al. 2000). The anticipated absence of polymorphisms in a subset of the relevant loci may reduce genetic variation and increase statistical power for the identification of genes (Shifman and Darvasi 2001).

Reproducibility of association studies. Many studies of common complex disorders have failed to replicate previously reported positive linkage and association findings. A meta-analysis of 370 association studies showed that significant between-study heterogeneity is frequent, and that the results of the first study correlate only modestly with subsequent research on the same association (Ioannidis et al. 2001). Both bias and genuine population diversity may explain why early association studies tend to over-estimate the disease predisposition conferred by a polymorphism. Discrepant associations may reflect different sets of genes acting in different samples or differences in the LD pattern between the populations studied. Alternatively, it may suggest that positive results may be artefactual, possibly due to the large number of markers tested or population stratification between cases and controls in an association study (Conneally 2003). Different studies may not have the same power to detect an underlying genetic effect due to differences in study design and in sample sizes used. The TDT design is robust with respect to population stratification, but it requires more individuals to be genotyped compared with the case-control study. Common errors in association studies include small sample sizes, subgroup analysis and multiple testing, random error, poorly matched control group, failure to attempt study replication, failure to detect LD with adjacent loci, over-interpreting results, positive publication bias and unwarranted candidate gene declaration after identifying association in arbitrary genetic region (Cardon and Bell 2001).

Parameters for reliable results in genetic association studies. Features of a reliable genetic association study include large sample sizes, small p-values, biologically meaningful associations and associations with alleles that affect the gene product in a physiologically meaningful way (Dahlman et al. 2002). Independent replication increases the confidence in an association finding. True association can be demonstrated both in family-based and population-based studies. It is essential to carry out power calculations, as a study with high statistical power is more reliable. A nominal p-value of 10^{-5} or 10^{-8} is recommended depending upon the nature of the study. For example, a genome-wide association study including 1000 000 tests requires a p-value in the range of 10^{-7} to 10^{-8} to provide a false positive rate of 0.05 (Risch and Merikangas 1996).

1.3.4 Genetic studies of Mendelian disorders

The last two decades have experienced a rapid achievement in mapping and cloning the genes involved in Mendelian disorders such as the X-linked muscular dystrophies (Koenig et al. 1987), cystic fibrosis (Kerem et al. 1989; Riordan et al. 1989; Rommens et al. 1989) and Huntington's disease (1993). These successes reflect the enormous power of linkage analysis when applied to Mendelian phenotypes (Risch 2000). Mendelian diseases are characterised by a one-to-one or near one-to-one correspondence between genotypes at a single locus and the observed phenotype. For Mendelian diseases, the total frequency of susceptibility mutations is generally very low (usually <1%). Mendelian disorders often have extremely high levels of allelic heterogeneity, which means that there are multiple different mutations (or alleles) in the same gene, all of which may lead to disease. For example, a British study of 424 families with haemophilia B identified 302 independent mutations in the factor IX gene (Green et al. 1999). The latter does not cause a problem in linkage analysis because all families will show linkage to the same chromosomal region. In effect, allelic heterogeneity provides the strongest evidence for a causal relationship between a gene and disease phenotype (Risch 2000). Recent evidence suggests that modifier genes and environmental factors contribute to some of the clinical variability observed in Mendelian diseases (Badano and Katsanis 2002). Modifier genes alter the form of the trait or its incidence in the presence of a major gene.

Finnish disease heritage. The term Finnish disease heritage is used to refer to the recessive diseases occurring in Finland at higher than usual frequencies compared with other European countries. At present, 35 diseases have been listed for Finnish disease heritage and the disease causing variant(s) have been identified in most of these disorders (Kere 2001).

1.3.5 Genetic studies of complex diseases

Encouraged by the success achieved in elucidation of the genetic aetiology of Mendelian traits, scientists began to turn their attention to common disorders, many of which have a major genetic component. The cause of common complex diseases is multifactorial with both genetic variation at multiple loci and environmental factors contributing to their development. Complex diseases may be very common, they have important consequences for the economics of medical provision.

The common disease / common variant hypothesis states that alleles at relatively high frequencies (>1%) represent a significant proportion of susceptibility alleles for common complex disease (Cardon and Bell 2001). Their high frequency implies that association studies in large population cohorts will be fruitful for identifying risk alleles. The key assumption of this model is that disease susceptibility is influenced by a few loci, each of which has a single major allele contributing to the phenotype. While the biological reality of these assumptions is unknown, they are critical to the success of association mapping experiments. A recent meta-analysis supports a contribution of common variants to susceptibility to common disease (Lohmueller et al. 2003), although this analysis may be biased due to the fact that associations with common alleles are easier to detect than associations with rare alleles. Conversely, it has been argued that rare alleles may equally contribute to the genetics of complex diseases even if their identification using the currently available techniques is difficult. The common disease / rare allele hypothesis states that there is no reason to expect that most genetic diseases results from common alleles. Population modelling predicts that disease loci exhibit extensive allelic heterogeneity in terms of the frequencies of the causative mutation (Pritchard 2001). The power of LD mapping of disease causing gene is greatly reduced if the causative mutation is rare. Simulation studies suggest that when the causative mutation is rare, most rare SNPs show no allelic association, but a small fraction show very strong association. In contrast, when the causative mutation is common, most rare SNPs show no allelic association, and a few show weak association, but almost all common SNPs show appreciable association.

Gene mapping of common complex disorders has been hampered by locus heterogeneity, epigenetic mechanisms, gene-environment and gene-gene interactions (epistasis). Locus heterogeneity denotes that mutations at multiple loci are involved in leading to the same phenotype. These mutations may act together or with environmental factors in additive or multiplicative way. Alternatively, a susceptibility locus of a major effect may act over a polygenic background of susceptibility genes. Other levels of genetic complexity include variable expressivity of the genes and pleiotropy, where the genes involved may have multiple effects. Epigenetic mechanisms include differential methylation pattern of GpC islands. In addition, there may be phenocopies, where the same clinical phenotype arises on the basis of environmental or random causes.

Penetrance refers to the proportion of people with the mutation who will manifest the disease. The power to detect association is drastically reduced by incomplete penetrance. A large number of subjects, whether as case-controls, families or affected sibling pairs, are needed for adequate power to detect linkage to a gene with only a partial effect on the phenotype. Especially in case of relatively rare complex diseases, international collaborations may be crucial in the recruitment of adequate number of affected families (Conneally 2003). Even for common complex disorders, collaborations can facilitate the identification of smaller gene effects. Collaborations among clinicians, epidemiologists, geneticists, mathematicians and bioinformatics experts will be needed to solve the genetic defects underlying complex disease.

A widely used approach for disease gene mapping of complex disease is initial whole genome screening using microsatellite markers with an average marker spacing of 10 cM. A two-stage approach, an initial and then a replication sample, is most commonly employed for mapping a complex trait. Subsequently, the broadly linked chromosomal regions are usually further refined with a denser set of microsatellite markers. The resolution of linkage studies is typically of the order of several cM, which in the human genome may correspond to several Mb of DNA, and 100s or 1000s of genes. Once the areas of interest are identified using linkage analysis, LD mapping is required to narrow further the region of the genome in which the disease gene must lie. Generally the chromosomal areas identified by the linkage studies are wide and a large number of SNP markers are required for fine mapping of the region. An alternative strategy is to identify candidate genes under the linkage peak and test the involvement of SNP markers in patients and healthy control individuals.

There are various alternative approaches for identification of important genes contributing to the complex disease phenotype. Once association is established, it is relevant to investigate whether any of the variants that were examined have a functional consequence. There may be biological, aetiological and pathological models of the disease of interest. Expression studies may provide important clues to the tissues and cells that are involved in the disease. Animal models of the disease of interest can provide important information on potential candidate genes and indicate relevant human homologues (Tabor et al. 2002). Mouse models, with their short generation times and high breeding efficiency, may provide shortcuts to disease gene identification, unequivocal proof that a particular mutation is causative (at least in mice), and rapid dissection of the molecular pathways in which the mutant protein acts (Peltonen and McKusick 2001).

1.3.6 Variation in the human genome

Initial analyses of the completed chromosomal sequences suggest that the number of human genes is lower than expected (about 35 000) (Lander et al. 2001). Major components of genetic complexity in the human genome may be variations in gene regulation and alternative splicing of gene transcripts explaining how one gene can have distinct functions in different tissues. Genome-wide analyses of alternative splicing indicate that 40-60% of human genes have alternatively spliced forms, and suggests that alternative splicing is one of the most significant components of the functional complexity of the human genome (Modrek and Lee 2002). Human genome is replete with several types of polymorphism: SNPs, repeat polymorphism and insertions and deletions, ranging from a single nucleotide to thousands of nucleotides in size. Most of the DNA sequence variation in the human genome is in the form of SNPs. Approximately 1 in 1,250 nucleotides is a SNP (Sachidanandam et al. 2001). They are present in coding and noncoding, as well as regulatory regions of the genes. SNPs are useful for population association studies because of their high density and mutational stability in the genome. The International SNP Map Working Group has provided a map of 1.4 million candidate SNPs across the genome (Sachidanandam et al. 2001). The biallelic nature of SNPs limits their informativeness. This can be overcome by constructing haplotypes of SNPs, which can make a region highly polymorphic. In some cases, determination of haplotypes or combinations of SNPs that are in LD may offer more power to detect associations than simply measuring individual SNPs. For disease mapping projects, it is desirable to study only those polymorphisms that affect the function of the protein or its expression, because these are most likely to result in a definable phenotype. As only a small proportion of the millions of sequence variations in our genomes are likely to have functional impact, identifying this subset of sequence variants will be one of the major challenges of the next decade.

Botstein and Risch characterised 27 027 mutations underlying Mendelian diseases. They noted that missense mutations resulting in in-frame amino acid substitutions were the most frequent type of mutation (59% of the total), whereas deletions account for 22% of all changes, insertions/duplications account for 7%, splice-site mutations account for 10% and mutations in the regulatory regions account for less than 1% of changes (Botstein and Risch 2003). They also noted that the severity of the disease tend to increase with the severity of the amino acid substitutions (Botstein and Risch 2003). A classic example of the correlation between severity of the phenotype and severity of the mutation is Duchenne (severe) and Becker (mild) muscular dystrophy: Duchenne is caused primarily nucleotide deletions leading to frame-shifts and Becker is caused by in-frame changes (Gillard et al. 1989). The clinical significance of mutation has also been demonstrated to correlate with the degree of cross-species evolutionary conservation (Botstein and Risch 2003).

1.3.7 Linkage disequilibrium and haplotype structure in the human genome

LD is the non-random association of alleles at adjacent loci. Factors that influence LD are recombination, mutation, genetic drift, population growth, admixture or migration, population structure, natural selection and gene conversion. Genetic drift describes the random fluctuation of allele and haplotype frequencies and these frequency changes are accentuated in small populations. Rapid population growth decreases LD by reducing genetic drift. Natural selection can affect the extent of disequilibrium by a hitchhiking effect, in which an entire haplotype that flanks a favoured variant can be rapidly swept to high frequency or even fixation. Natural selection can also affect LD through epistatic selection for combinations of alleles at two or more loci on the same chromosome. In a gene conversion event, a short stretch of one copy of a chromosome is transferred to another during meiosis. Rates of gene conversion in humans are high and are important in LD between very tightly linked markers (Ardlie et al. 2001).

If LD mapping strategy is to be applied successfully in disease gene mapping, it is essential to know how far usable levels of disequilibrium extends in the human genome and how much this varies from one population to another. Abecasis et al. estimated that physical distance could account for less than 50% of the variation in LD. The remaining variation was probably due to a combination of drift, demographic factors, selection and variable rates of mutation, recombination and gene conversion (Abecasis et al. 2001). Eaves et al. and Taillon-Miller et al. noted substantial variability in disequilibrium levels beyond that which can be explained by differences in distances and allele frequencies (Eaves et al. 2000; Taillon-Miller et al. 2000).

Empirical evidence suggests that the genome is broken down into blocks of strong haplotype structure, characterised by low haplotype diversity, strong associations between alleles and rare recombination, separated by shorter regions of shattered haplotype structure, characterised by high haplotype diversity, weak allelic associations and multiple recombination events (Daly et al. 2001; Phillips et al. 2003). Within high LD regions, allelic dependence yields redundancy among markers and improves the chances of detecting association using only a fraction of all the known markers (Johnson et al. 2001). The Haplotype Mapping (HapMap) project is in progress to identify and catalogue common haplotypes and their characteristic SNPs across the genome for major races. HapMap will undoubtedly benefit the mapping of regions of interest in complex diseases and lay a foundation for the genome-wide association studies (Judson et al. 2002). There is a concern that association analysis based on identification of common haplotype blocks will perform poorly if the allele frequency is as low as 0.1 (Pritchard and Cox 2002). However, rare variants are likely to be young, and hence may lie in conserved haplotypes that extend across several haplotype blocks.

In conclusion, the reconstruction of accurate population specific linkage and haplotype maps is likely to be of key importance for the identification of causative polymorphisms underlying complex traits. Well-designed and large studies are likely to yield successful results in future gene mapping projects of complex diseases.

1.4 AIMS OF THIS PROJECT

AS is a complex and highly heritable disorder exhibiting a strong association with *HLA-B27*. However, *HLA-B27* is estimated to account for 16-50% of the total genetic variation of the disease leaving a large component of the genetic variance to be discovered. This project aims to further define the HLA region and investigate the contribution of candidate genes outside the HLA region in AS.

The specific aims are:

1. To investigate the effect of polymorphic variants near and within the *TGFB1* gene in susceptibility to, and severity of, AS in an English and Finnish population
2. To develop a novel *HLA-DRB1* genotyping method based on primer extension reactions
3. To investigate the effect of HLA genes and haplotypes in susceptibility to, and clinical characteristics of, AS in a Finnish population
4. To investigate the effect of the *CYP2D6* gene in susceptibility to AS in a Finnish population

CHAPTER 2 - SUBJECTS AND METHODS

2.1 AS CASES AND CONTROLS

2.1.1 Case definition

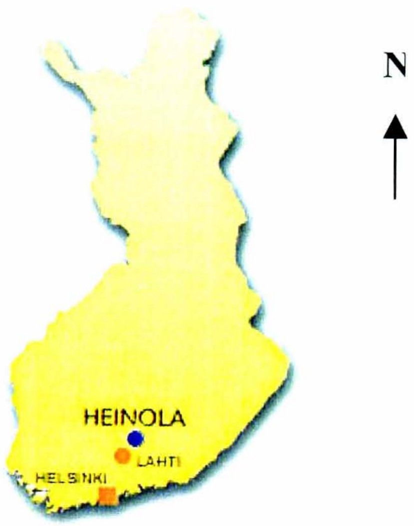
Cases with AS defined by the modified New York diagnostic criteria (van der Linden et al. 1984a) were selected for the study. The sacroiliitis was confirmed by a qualified radiologist and the diagnosis of AS was confirmed by a qualified rheumatologist.

2.1.2 Finnish families

Finland is a northern European country covering 337 030 km^2 with a total population size of 5.2 million people. All the Finnish AS patients were recruited from the Rheumatism Foundation Hospital in Heinola, which is located in southern Finland 133 kilometres north of Helsinki (figure 2.1). It serves as a specialist unit responsible for a population of 200 000 people and a rheumatologic rehabilitation centre nationwide. A structured questionnaire was used to assess the presence of AAU, IBD, psoriasis and peripheral arthritis, age of symptom onset, age at diagnosis, disease duration and disease severity scores including the Bath AS Disease Activity Index (BASDAI) (Garrett et al. 1994), Bath AS Functional Index (BASFI) (Calin et al. 1994) and Bath AS Global Score (BAS-G) (Jones et al. 1996) (see Appendix 1). The original English language questionnaire was translated into Finnish, and then validated by reverse translation. At the time of completing this thesis, sample recruitment was still continuing. At the time of conducting the *CYP 2D6* and *TGFB1* study, we had 437 individuals from 170 families available. For the HLA studies, 673 individuals from 261 families were available.

Figure 2.1 Map of Finland and location of Heinola.

The figure is obtained from www.heinola.fi.



2.1.3 British families

One thousand three hundred and thirty-two individuals from 212 parent-case trio families and 184 AS affected-sibling pair families were recruited from the UK for the studies reported in this thesis. AS patients were identified from several sources: the Royal National Hospital for Rheumatic Disease AS database; patients attending the Nuffield Orthopaedic Centre, Oxford; in response to public appeals; and by referral from British rheumatologists. Among the parent-case trios there were 31 parent-case affected pairs; other families were simplex i.e. one affected subject in family.

2.2 LABORATORY METHODS

2.2.1 DNA extraction

Thirty-six mls of blood per individual was collected in sodium EDTA anticoagulation tubes and stored at -20 °C. DNA was extracted from the white blood cells using the guanidinium hydrochloride (GuHCl) DNA extraction method.

Defrosted blood was poured into two 50 ml Falcon tubes, 25 ml of Lysis Buffer (see Appendix 2 for composition) was added and the solution was mixed by inversion. The mixture was centrifuged at 2500 rpm (1467 rcf) (Heraeus Sepatech Megafuge 2.0R) for 10 min to pellet the nuclei of the white cells. The supernatant was decanted. Twenty mls of Lysis buffer was added and the solution was vortexed until the pellet was completely dispersed. Centrifugation at 2500 rpm (1467 rcf) for 10 min was performed and supernatant decanted.

Three and half mls of 6 M GuHCl was added to lyse the nuclei. The solution was vortexed to resuspend the pellet. Two hundred and fifty µl of 7.5 M ammonium acetate was added. Twenty-five µl of 20 mg/ml Proteinase K solution was added to digest the protein. Two hundred and fifty µl of 10% SDS was added. Incubation at 60°C was performed for one hour. The tubes were shaken after ten minutes of incubation in order to disperse clumps of SDS.

Two ml of $CHCl_3$ was added and the tubes were thoroughly mixed by inversion. Centrifugation at 2500 rpm (1467 rcf) was performed for 5 min. The upper layer was collected into a new 50 ml Falcon tubes. Fifteen mls of 100% ethanol was added and mixed thoroughly by inversion to precipitate the DNA. The tubes were centrifuged at 2500 rpm (1467 rcf) for 2 min to pellet the DNA. The supernatant was decanted and 10 ml of 80% ethanol was added. The precipitated DNA was hooked out with a tip and removed into a clean Eppendorf tube. Eight hundred µl of 80% ethanol was added, the tubes were centrifuged at 13000 rpm (15115 rcf) in a microfuge (IEC Micromax) and the ethanol was decanted. The tubes were centrifuged again at 13000 rpm (15115 rcf) in a microfuge and the last visible traces of ethanol were removed with a pipette. The Eppendorf tubes were left open for 10-20 min to dry the DNA pellets. One ml of 1xTE buffer (see Appendix 2 for composition) was added the tubes were left overnight at room temperature to resuspend. Following extraction, undiluted DNA samples were stored at -70°C. Samples to be used in genotyping were diluted with sterile dH_20 to a concentration of 20 ng/µl and stored at -20°C until required.

2.2.2 DNA quantification

Nucleotides in solution absorb light in the ultraviolet (UV) region of the spectrum, with the maximum absorption at a wavelength of 260 nm. The maximum absorbance of proteins in the UV range is at 280 nm. Nucleic acids can therefore be quantified and the purity of a preparation can be estimated by measuring the absorption at both 260 nm and 280 nm.

The absorbance measurements at wavelengths of 260 nm and 280 nm were performed to determine the amount and purity of DNA using Beckman DU® 640 spectrophotometer. The term optical density is commonly used instead of absorbance. An optical density (OD) of 1.0 at 260 nm is equivalent to approximately 50µg/ml so DNA concentration was calculated as follows:

DNA concentration µg/ml = 50 x dilution x OD_{260}.

The purity of DNA was assessed by measuring the ratio of OD_{260}/OD_{280}. If protein is present in the solution, the OD_{280} increases more than the OD_{260}, so the ratio OD_{260}/OD_{280} decreases. Pure DNA has a ratio of OD_{260}/OD_{280} of 1.8. Samples with the OD_{260}/OD_{280} ratio <1.5 were re-extracted with phenol and $CHCl_3$ (to remove any phenol that has entered the aqueous DNA containing phase) and precipitated again.

Some of the DNA concentrations were determined using the PicoGreen™ kit (Molecular Probes, Leiden, Netherlands). The PicoGreen™ double-stranded DNA (dsDNA) quantitation reagent becomes intensely fluorescent upon binding nucleic acids, allowing selective detection of dsDNA in solution (Singer et al. 1997). The PicoGreen™ solution and the stock DNA samples were diluted by a factor of 1 in 200 in TE buffer. Fifty µl of the diluted DNA and 50 µl of diluted PicoGreen™ solution were added in a MicroAmp Optical 96-well plate. A set of 8 standards ranging from 0 to 1000 ng/ml were included in each plate. The fluorescence emitted by each sample was measured using an ABI Prism 7200 Sequence Detector. A standard curve of concentration against fluorescence was created using the standards and the concentration of each sample determined using this curve. The PicoGreen™ dsDNA quantitation reagent allows a sensitive, specific and rapid means to detect and quantitate DNA.

2.2.3 Polymerase Chain Reaction (PCR)

PCR amplifies a specific segment of DNA between two flanking oligonucleotide primers. Oligonucleotide primers are generally 15-30 nucleotides long and designed to hybridise specific sequences (5' and 3') flanking the region of interest in the target DNA. In a PCR reaction three steps, denaturation at 94 °C, annealing at 50-60 °C and extension at 72 °C are repeated a number of times (generally about 30 times) in order to achieve the required amplification. DNA polymerase extends the primers after they anneal to the template, creating a duplicated complementary strand. Heat stable DNA polymerase such as Taq (derived from *Thermus aquaticus*) DNA polymerase is stable at the denaturing temperature of DNA and is suitable for PCR amplification.

Magnesium ions are necessary to deoxynucleotide triphosphate (dNTP) incorporation and stabilise primer annealing. The concentration of $MgCl_2$ can have a substantial effect on the specificity and yield of the amplification. Generally, excess magnesium ions will result in the accumulation of non-specific amplification products and insufficient magnesium ions will reduce the yield. $MgCl_2$ concentration is therefore optimised for each primer pair.

Primer design. There are computer programs available to aid the oligonucleotide design. PRIMER3 [www.genome.wi.mit.edu/cgi-bin/primer/primer3 www.cgi] designs suitable PCR primers given the stretch of nucleotide sequence. The OLIGO [www.oligo.net] program analyses given oligonucleotides. PCR primers were designed to avoid stable hairpin loop structures. The primers were designed to have a G + C content of 40 – 60% and not to have areas of base complimentarity either within a primer or between pairs of primers. In particular, overlaps at the 3' ends were avoided to reduce the incidence of an artifactual product, called 'primer-dimer'. Complementary 5' extensions may be added to primers to allow a variety of useful post-amplification manipulations. Optimal primer concentrations were determined empirically for each PCR reaction. Specific concentrations and volumes used in individual mastermixes as well as cycling conditions used are detailed in the methods section of each chapter.

2.2.4 Agarose gel electrophoresis

DNA molecules can be separated according to their molecular weight using agarose gel electrophoresis. Agarose is a linear polymer composed of alternating residues of D- and L-galactose. Negatively charged DNA migrates toward the positively charged electrode. Smaller fragments of DNA migrate faster and therefore, fragments of different sizes can be separated. Two percent agarose gels were made using 12 g of agarose (Sigma Aldrich, Steinheim, Germany) and 600 ml of 1 x TBE buffer (see Appendix 2 for composition).

Nucleic acids that have been subjected to electrophoresis through agarose gels may be detected by staining and visualised by illumination with 300 nm UV light. Fluorescent dye 2,7-diamino-10-ethyl-9-phenylphenanthridinium bromide (ethidium bromide) was used to stain the agarose gels. Ethidium bromide is a small molecule that inserts between the nucleotides of DNA and strongly fluoresces under UV illumination. When ethidium bromide binds to DNA, it is effectively concentrated and its fluorescence increases, so the nucleic acid shows up as a bright band on a dim background of unbound ethidium bromide. PCR products were visualised under UV light on an AlphaImager (Flowgen, Staffordshire, UK).

2.2.5 SNP ascertainment

2.2.5.1 Amplification Refractory Mutation System (ARMS)

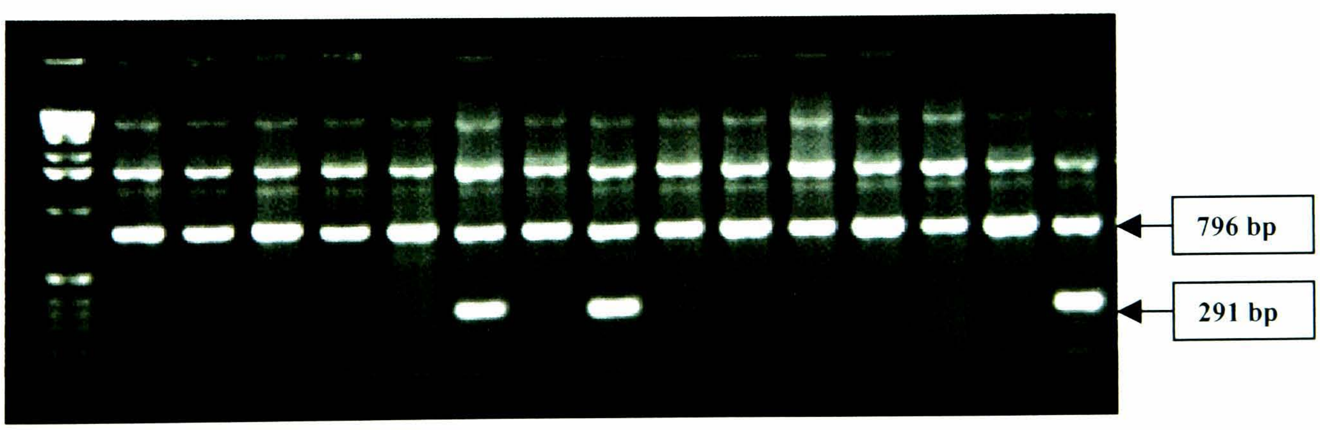
ARMS was described in 1989 as a PCR-based method to detect any known SNPs in genomic DNA (Newton et al. 1989) (figure 2.2). The basis of this method is that oligonucleotides with a mismatched 3'-residue will not function as primers in the PCR under appropriate conditions. The 3'-mismatch principle can be used to identify virtually any SNP mutation within a PCR reaction. Identification of the alleles is based on the presence or absence of amplified product observed after agarose gel electrophoresis.

Figure 2.2 Agarose gel view of ARMS genotyping of *TGFB1* -800-509 haplotype 2/1.

Control primers amplifying a 796 bp fragment from the third intron of *HLA-DRB1* gene were included in all the reactions. Allele specific fragment is 291 bp. From left to right: Lane 1 – 1 kb ladder; Lanes 2-6, 8, 10-15 – *TGFB1* -800-509 haplotype 2/1 negative; Lanes 7, 9, 16 – *TGFB1* -800-509 haplotype 2/1 positive samples.

Lanes:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16



2.2.5.2 PCR-Restriction Fragment Length Polymorphism (PCR-RFLP)

PCR-RFLP method relies on the initial amplification of the fragment of interest by PCR and subsequent digestion of fragment with a suitable restriction endonuclease which cuts at specific restriction sites. Over 200 different sequence specificities of restriction endonucleases have been characterised which has made it possible to genotype many SNPs by this method.

2.2.5.3 Primer extension method

Primer extension reactions were performed using an ABI Prism® SNaPshot™ technology (Applied Biosystems, Foster City, CA, USA). The primer extension reaction requires a prior PCR amplification of the fragment of interest and subsequent purification of the PCR product. Shrimp Alkaline Phosphatase (SAP) was used to remove unincorporated dNTPs and Exonuclease I (Exo I) was used to remove unincorporated primers. Primer extension method uses a primer that anneals immediately adjacent to the SNP and is based on the extension of the primer using fluorescent dideoxynucleotide triphosphates (ddNTP) that are complimentary to the SNP. Unlike normal dNTPs, ddNTPs cannot be extended by DNA polymerase which results in termination of the reaction. The subsequent extension product is analysed on an ABI Prism 3700 capillary electrophoresis instrument.

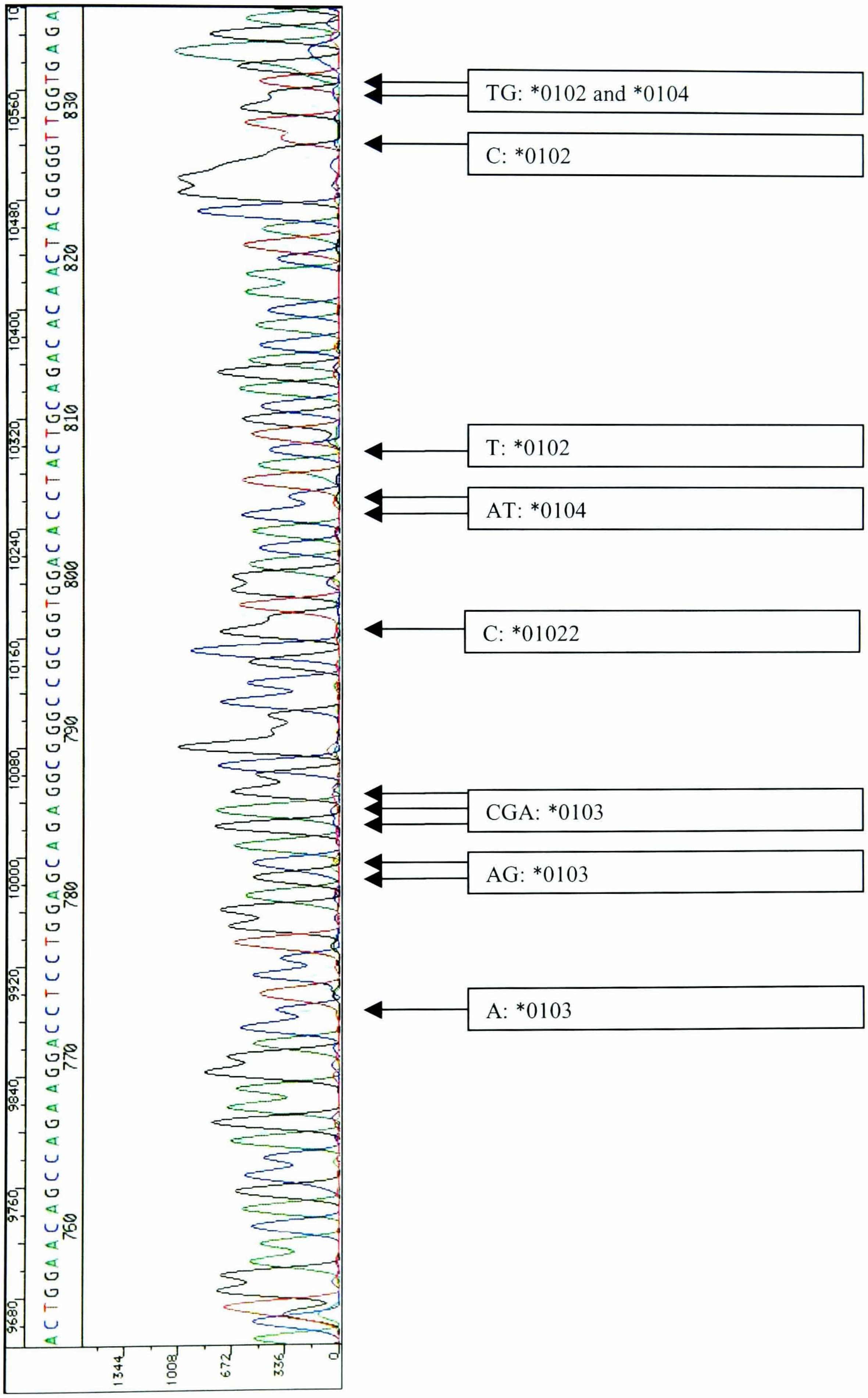
2.2.6 DNA sequencing

Two automated methods, dye-primer and dye-terminator, are commonly used for DNA sequencing. They are based on Sanger dideoxy DNA sequencing method (Sanger et al. 1977). In sequencing with dye-labeled terminators, the dyes are attached to the terminating ddNTP (Lee et al. 1992; Rosenblum et al. 1997). The advantage of dye-terminator chemistry is that four chain-extension reactions can be carried out with the same primer in a single reaction and only properly terminated products are detected. In sequencing with dye primers, the fluorescent dyes are attached to the 5' end of the primer oligonucleotide (Smith et al. 1985; Smith et al. 1986; Voss et al. 1989). Four separate reactions are carried out for each DNA sample, each reaction containing a different dye-labeled primer. This set of four reactions is then mixed and loaded into a single channel for gel electrophoresis. As the sequence undergoes electrophoresis in an automated sequencer, a laser beam excites the dyes. The fluorescent signal can be reproduced in the form of an intensity profile. Yellow, red, green and blue colour represents guanine (G), thymine (T), adenine (A) and cytosine (C) bases, respectively. The advantage of dye-primer sequencing is that the base-calling is more accurate, especially with heterozygote samples. The disadvantage of the dye primer method is the requirement of for four separate extension reactions and four dye-labeled primers for each template (Sambrook and Russell 2001).

HLA-DRB1 gene sequencing was performed using the dye-primer sequencing (figure 2.3). Specific concentrations, volumes and cycling conditions used in sequencing reactions are presented in the methods section of chapter 5.

Figure 2.3 An electropherogram view of *HLA-DRB1*0101* positive sample.

The sequence spans the codons 89-117 of the *HLA-DRB1* gene (reference sequence ID: OTTHUMG00006006904). The arrows point to the location of the polymorphisms which define this as an *HLA-DRB1*0101* subtype sequence. The corresponding bases found in the other *HLA-DRB1*01* subtype sequences are denoted in the boxes.



2.3 ANALYTICAL METHODS

2.3.1 Genotype error checking

The GAS program version 2.0 (Alan Young, Oxford University, 1995) was used to check for Mendelian inheritance of alleles. The program MERLIN was used for error detection (Abecasis et al. 2002) (http://bioinformatics.well.ox.ac.uk/Merlin). MERLIN-ERROR detects genotypes that cause unlikely recombination events between tightly linked SNP markers suggestive of genotyping errors.

2.3.2 Testing for Hardy-Weinberg Equilibrium (HWE)

The HWE is named after the English mathematician G. H. Hardy (1877-1947) and the German physiologist Wilhelm Weinberg (1862-1937), who in 1908, independently formulated the principle and deduced its theoretical predictions of genotype frequency.

In the Hardy-Weinberg model, the mathematical relation between the allele frequencies and the genotype frequencies is defined by

$p^2 + 2pq + q^2 = 1$

$p + q = 1$

in which p^2, $2pq$ and q^2 are the frequencies of the genotypes AA, Aa and aa in zygotes and p and q are the allele frequencies of A and a.

The assumptions of HWE are diploid organism, sexual reproduction, non-overlapping generations, identical allele frequencies in males and females, random mating, very large (in theory infinite) population size, negligible migration, no mutation and no natural selection (Hartl and Clark 1997). Deviation from HWE can occur due to a major violation of one of the HWE assumptions, sampling bias or genotyping error. The chi-squared test was used to examine whether the genotype frequencies were in HWE. P-values >0.05 were accepted as nonsignificant deviation from HWE.

2.3.3 Haplotype reconstruction

Traditionally haplotypes are reconstructed by using the information from the available family members. Families are not always available and a number of approaches have been developed to reconstruct haplotypes without family data. Haplotypes can be reconstructed from the genotype data obtained from the unrelated individuals using a maximum parsimony method (Clark 1990), the expectation-maximization (EM) algorithm (Excoffier and Slatkin 1995) or a coalescent model-based method using the program PHASE (Stephens et al. 2001). In the *TGFB1* data set PHAMILY [Hans Ackerman thesis, Oxford University, 2001] and PHASE (version 1.0) were used for haplotype reconstruction for the sporadic AS cases and parent case trio families (Stephens et al. 2001). The *HLA-DRB1-B27* haplotypes were reconstructed using a family based haplotyping program, SIMWALK2 (Sobel and Lange 1996).

PHASE reconstructs haplotypes using a Bayesian method of genetic modelling. Specifically, it uses Gibbs sampling, a type of Markov chain-Monte Carlo algorithm, to obtain an approximate sample from the posterior distribution of haplotypes (H) given genotypes (G), Pr (H|G). The algorithm starts with an initial guess haplotype, repeatedly chooses an individual at random, and estimates that individual's haplotypes under the assumption that all other haplotypes are correctly reconstructed. Repeating this process enough times results in an approximate sample from Pr (H|G). It also estimates the uncertainty of each phase call to prevent overconfidence in statistically reconstructed haplotypes. PHASE has been demonstrated to be more accurate than the EM algorithm (Excoffier and Slatkin 1995) or a maximum parsimony method (Clark 1990). The accuracy of PHASE and EM algorithm is improved with increasing sample size. The disadvantage with PHASE program and EM algorithm is that, without pedigree information, rare haplotypes are occasionally miscalled when their presence/absence must be inferred (Tishkoff et al. 2000).

PHAMILY assesses family haplotype data where phase can be directly established, which can subsequently be incorporated into the PHASE program. PHAMILY reads data from parent-case trio families and assigns gametic phase to unambiguous positions. PHAMILY provides two output files: one containing the partially reconstructed parental haplotypes and the second indicating whether phase is ambiguous or unambiguous at each site. The second output file can be used as an additional input file for PHASE to increase the accuracy of haplotype reconstruction by using all available information.

HLA haplotypes in pedigrees were reconstructed using an algorithm based on genetic descent graphs, implemented in the SIMWALK2 software (Sobel and Lange 1996). SIMWALK2 is based on a Markov Chain Monte Carlo procedure and inheritance graphs. SIMWALK2 performs stochastic simulation on pedigrees and estimates the single most likely configuration of haplotypes. SIMWALK2 assumes that the loci are in linkage equilibrium.

2.3.4 Linkage analysis

2.3.4.1 Nonparametric linkage analysis

Nonparametric multipoint linkage analysis was performed using GENEHUNTER-PLUS (Kong and Cox 1997), a modification of GENEHUNTER (Kruglyak et al. 1996). GENEHUNTER calculates a NPL statistic from the sharing of alleles ibd among affected individuals within a pedigree (Kruglyak et al. 1996). The NPL method can analyse the data using a pairwise approach (using the NPL_{pairs} statistic) by measuring the number of alleles shared ibd by a pair of affected relatives. It has also been extended to provide simultaneous comparison of alleles in all affected individuals in a pedigree (the NPL_{all} statistic). NPL_{all} calculates the observed ibd probability over all configurations and puts extra weight on families in which three or more affected individuals share the same allele. NPL_{all} statistic is most suitable for analysis of dominant loci, since more affected individuals per family are likely to share the same allele with this type of inheritance. The NPL approach is inherently multipoint, since it calculates the ibd probability for any given data point along the chromosome, using all available marker data on the chromosome. Because the true inheritance pattern is not always known with certainty but significance is evaluated using the null distribution expected in the case of complete informativeness, the NPL score will tend to be conservative. The level of conservativeness depends on the marker heterozygosity, the number of untyped individuals and marker spacing.

To resolve the conservativeness of the NPL method when the descent information is incomplete, the GENEHUNTER-PLUS software uses the modified NPL_{lr} statistic that is well approximated by a normal distribution (Kong and Cox 1997). GENEHUNTER-PLUS provides an accurate likelihood-ratio test to assess evidence of linkage and also allows the calculation of a nonparametric LOD score that is based on allele sharing, and unlike the NPL score, can be interpreted in the same way as a traditional LOD score.

2.3.4.2 Quantitative linkage analysis

The Haseman-Elston method consists of regressing the squared sib-pair quantitative trait phenotype difference on the proportion of alleles that the sibs are estimated to share ibd at the marker locus (Haseman and Elston 1972). A negative slope suggests linkage because it correlates similarity at a trait locus with similarity at a marker locus (i.e. as the sib-pair trait difference decreases, their genetic identity increases). A modified Haseman-Elston algorithm (Sham et al. 2002) that is adapted for use in general pedigrees was employed to perform the quantitative linkage analysis. This method is implemented by the program MERLIN-REGRESS (Abecasis et al. 2002). The basis of this approach is to test whether relatives with close phenotypic values share more ibd alleles than relatives with more distant values. The method takes into account that ibd information may be incomplete, as is often the case with the real data. MERLIN-REGRESS requires specification of the trait distribution parameters (mean, variance and heritability) in the general population. The method is limited to additive genetic effects on a single quantitative trait.

2.3.5 Association analysis

2.3.5.1 Transmission disequilibrium test (TDT)

The TDT test was used as a within-family association method to assess the association between the polymorphisms and disease susceptibility (Spielman et al. 1993). The concept that underlies TDT is that, in the presence of association between a genetic marker and disease susceptibility locus, the probability of transmission of a marker gene from parents to an affected offspring is increased from the 0.5 value predicted by Mendelian inheritance. The statistic of TDT is

$$\chi^2_{TDT} = \frac{(a - b)^2}{a + b}$$

in which a is the number of transmitted alleles from heterozygous parents and b is the number of non-transmitted alleles from heterozygous parents. This has a chi-squared distribution with one degree of freedom. Untransmitted alleles act as an internal control for transmitted alleles, and therefore, TDT test is robust to population stratification, which may cause spurious associations in case-control studies. However, TDT only counts heterozygous parents in the analysis and requires more genotyping than case-control studies. Thus, the false negative rate of TDT is higher than that of the case-control study, which causes loss of power.

The program TRANSMIT version 2.5 uses a new approach to transmission disequilibrium testing based on the score test (Clayton 1999). TRANSMIT can deal with situations in which transmission is uncertain. Such situations arise when transmission of a multilocus marker haplotype is considered, since haplotype phase is often unknown. For single-locus markers, transmission is uncertain if one or both parents are missing. By use of the robust variance estimator, more than one affected offspring per family may be used to assess association independent of linkage. P-values can be obtained by statistical simulation using 1000 replicates.

2.3.5.2 Quantitative transmission disequilibrium test (QTDT)

Associations of the polymorphisms and haplotypes with quantitative data such as the age of disease onset, BASDAI and BASFI were assessed using the QTDT program (Abecasis et al. 2000). QTDT incorporates variance components methodology in the analysis of family data and partitions association into between- and within- family components. The program includes a permutation framework, which allows estimation of p-values. This corrects for small sample sizes and deviations of quantitative traits from multivariate normality. One thousand simulations were performed for all the polymorphisms. Disease duration and gender correlated with BASDAI and BASFI in this data set and were used as covariates in the analysis.

2.3.5.3 Case-control analysis

Case-control study design provides an alternative to the TDT analysis at the population level. Case-control studies compare the allele frequencies at loci among disease and control populations. The disease susceptibility mutations will exhibit statistically significant differences in frequency between cases and controls. The statistical significance of the case-control analysis is assessed by the chi-squared test, which compares the observed numbers in each of the four categories in the contingency table with the numbers to be expected (table 2.1).

Table 2.1 Observed a) and expected b) number of alleles in a 2 x 2 contingency table.

a)

OBSERVED	DISEASE +	DISEASE -	
MARKER +	A	C	A+C
MARKER -	B	D	B+D
	A+B	C+D	A+B+C+D

b)

EXPECTED	DISEASE +	DISEASE -
MARKER +	((A+C)/(A+B+C+D))*(A+B)	((A+C)/(A+B+C+D))*(C+D)
MARKER -	((B+D)/(A+B+C+D))*(A+B)	((B+D)/(A+B+C+D))*(C+D)

$$\text{Chi-squared test} = \frac{\Sigma\,(\text{observed} - \text{expected})^2}{\text{expected}} \quad \text{with 1 degree of freedom.}$$

The relative risk (RR) of disease in marker carriers compared with non-carriers was calculated as follows.

$$RR = \frac{A/(A+C)}{B/(B+D)}$$

The odds ratio (OR) was calculated by the cross-product ratio (Kirkwood 1988). OR=AD/BC. In the HLA case-control study, the Haldane's modification of Woolf's method was used to calculate the OR as this method avoids problems when entries are zeros (Svejgaard and Ryder 1994).

$$OR = \frac{(A + \frac{1}{2})(D + \frac{1}{2})}{(B + \frac{1}{2})(C + \frac{1}{2})}$$

2.3.6 Linkage disequilibrium calculation

The pairwise LD between the SNPs was calculated using the Lewontin's standardised disequilibrium coefficient D' (Lewontin 1988). D quantifies disequilibrium as the difference between the observed frequency of a two-locus haplotype and the frequency it would be expected to show if the alleles were segregating at random. If f_{AB} represents the observed frequency of gamete AB, and f_A and f_B represent the frequencies of alleles A and B, the LD is calculated as follows:

$$D_{AB} = f_{AB} - f_A \times f_B$$

This statistic is strongly influenced by the magnitude of f_A and f_B independently of the effects of LD. To overcome this bias, normalised statistic was used D':

$$D' = D_{AB} / D_{max} \text{ where}$$

$$D_{max} = \min [f_A f_B,(1-f_A)(1-f_B)] \text{ when } D_{max}<0 \text{ or}$$

$$D_{max} = \min [f_A(1-f_B),(1-f_A)f_B] \text{ when } D_{max}>0$$

This statistic measures the magnitude of LD on a scale from –1 to 1 and is less dependent on allele frequencies than the non-standardised formula. The statistical significance of the finding was assessed using the chi-squared test.

Pairwise tests of LD were performed using the HAPLOXT program, which uses definitions that have been expanded to allow for testing of multiallelic markers (Abecasis and Cookson 2000). HAPLOXT program uses Hedrick's modification of Lewontin's standardised disequilibrium coefficient D' (Hedrick 1987), which uses the absolute values of LD and thus the range of LD values fluctuates between 0 and 1. A graphical summary of the LD was produced using the software package GOLD

(Abecasis and Cookson 2000), which provides graphical output of both the LD measured as D' and the associated p-values.

2.3.7 Power calculation

The power of the TDT was determined using TDT Power Calculator as described (Chen and Deng 2001). This program does not allow for missing parents in the data, and thus in this study gives an overestimate of the actual power. Risch and Merikangas demonstrated that TDT has much more power than linkage using affected sibling pairs even if much more stringent type I error is applied to TDT assuming 100% LD between the marker and the disease associated allele (Risch and Merikangas 1996). Factors that influence power of the TDT study are disease allele frequencies, family structure, sample size, LD between the marker allele and the disease locus, penetrance of the disease, mode of inheritance, disease prevalence and type I error rate.

The power of the case-control studies was determined by the Genetic Power Calculator (http://statgen.iop.kcl.ac.uk/gpc/) (Purcell et al. 2003). It is a website for performing power calculations for the design of linkage and association genetic mapping studies of complex traits. It implements variance components equations and others for power calculations. For tests of association, Genetic Power Calculator uses the variance components test described by Fulker (Fulker et al. 1999).

2.3.8 Entropy maximization

The program Entropy was used to measure the haplotypic diversity of the SNPs used www.well.ox.ac.uk/~rmott/SNPS]. Entropy (E) was defined as

$$E = -\sum_{i=1}^{\kappa} p_i \log_2 p_i$$

and $p_i = f_i/2n$

in which f_i is the frequency of the ith haplotype, n is the number of individuals genotyped and κ is the number of unique haplotypes in the sample. Entropy can be used for haplotype tagging, the process of selecting a subset of SNPs that best approximate the haplotypic diversity in the population. Haplotypes constructed using SNPs and the haplotype population frequencies are used as an input for Entropy.

2.3.9 PipMaker program

The program PipMaker (http://bio.cse.psu.edu/pipmaker/) was used to compare the coding and non-coding sequences between human and mouse (Schwartz et al. 2000). PipMaker is a website for comparing two long DNA sequences to identify conserved segments. PipMaker produces a percent identity plot (pip), which shows both the position in one sequence and the degree of similarity for each aligning segment between the two species. It also provides a plot of the position of each aligning segment in both species. PipMaker requires two sequence files to be compared and a file containing the positions of known or predicted exons so that the exons can be marked in the pip. RepeatMasker (Smit, AFA & Green, P, http://ftp.genome.washington.edu/RM/RepeatMasker.html) is used to generate the coordinates of interspersed repeats in the first sequence in order to avoid uninformative and time-consuming alignments among repeats. For the pip the program plots the position and percent identity of each gap-free segment of the alignments. The top horizontal axis is automatically annotated with the positions from the repeats and exons. The positions of the CpG islands are also computed and displayed along the top axis.

CHAPTER 3 - THE EFFECT OF TRANSFORMING GROWTH FACTOR β1 GENE POLYMORPHISMS IN AS

This chapter presents the results of a study designed to determine whether genetic polymorphisms in or near the transforming growth factor β1 (*TGFB1*, OMIM reference number 190180) locus are associated with susceptibility to, or severity of, AS. The theories behind association and linkage analysis and haplotype reconstruction are presented in chapters 1 and 2.

3.1 INTRODUCTION

AS exhibits a strong genetic component determining both susceptibility to and severity of the disease; except for *HLA-B27*, the genes responsible have not been definitely identified (Brown et al. 1997b; Hamersma et al. 2001). Non-MHC genes are estimated to account for at least half of the genetic variance of AS susceptibility (Brown et al. 2000b). The male:female ratio for AS is 2.5:1 – 5:1 (Kennedy et al. 1993), but the gender bias is not explained by the linkage to the X chromosome (Hoyle et al. 2000). The gene (*TGFB1*) for transforming growth factor (TGF) β1 is located on chromosome 19q13.2, where suggested evidence of linkage was reported in both the whole genome screen previously performed by Laval et al (Laval et al. 2001) and the North American Spondylitis Consortium genome screen (Reveille 2002). The peak of linkage on chromosome 19 was observed with the marker D19S420 (LOD score 3.58) (Laval et al. 2001), which is located 1.8 Mb from the *TGFB1* gene. Strong evidence of linkage to chromosome 19 was also identified in a genome-wide scan of IBD (Rioux et al. 2000), a disease that is clinically associated with spondyloarthritis.

3.1.1 TGF-β1 cytokine family

TGF-β1 is a multifunctional cytokine that is capable of inducing a spectrum of cellular events including cell proliferation, cell cycle arrest, differentiation, matrix deposition, immune response and programmed cell death (Kaartinen and Warburton 2003). TGF-β1 was initially identified by its ability to reversibly induce phenotypic transformation of fibroblast cell lines (Roberts et al. 1981). It belongs to a large family of secreted growth factors, called TGF-β superfamily, which contains different TGF-β isoforms, inhibins/activins, bone morphogenetic proteins and several other secreted factors. Mammalian species express three distinct isoforms of TGF-β, TGF-β1, -β2 and -β3, encoded by unique genes located on different chromosomes (human chromosomes 19q13, 1q41 and 14q24, respectively). Each isoform is expressed in both a tissue-specific and a developmentally regulated fashion (Blobe et al. 2000). TGF-β1 is the predominant isoform in most cells and tissues. TGF-β1 is most abundant in platelets, bone and spleen. It is also present in the circulation covalently bound to α_2–macroglobulin suggesting an endocrine function of this isoform. In addition significant quantities of TGF-β1 are sequestered in extracellular matrix. TGF-β2 is mostly present in fluids such as the aqueous and vitreous humours of the eye and amniotic fluid.

3.1.2 Structure and polymorphisms of *TGFB1* gene

The structure of the *TGFB1* gene has been characterised and is presented schematically in figure 3.1. The *TGFB1* gene consists of seven exons encompassing 23.5 kb. The genomic region around the *TGFB1* gene is presented in figure 3.2.

Figure 3.1 Position of the *TGFB1* gene on chromosome 19 and diagram of the *TGFB1* locus drawn to scale.

The yellow boxes illustrate seven exons of the gene (numbered in blue) and the arrows point to the location of the five polymorphisms genotyped in this study within and in the promoter region of the *TGFB1* gene. The figure of chromosome 19 was obtained from the Ensembl webserver at www.ensembl.org.

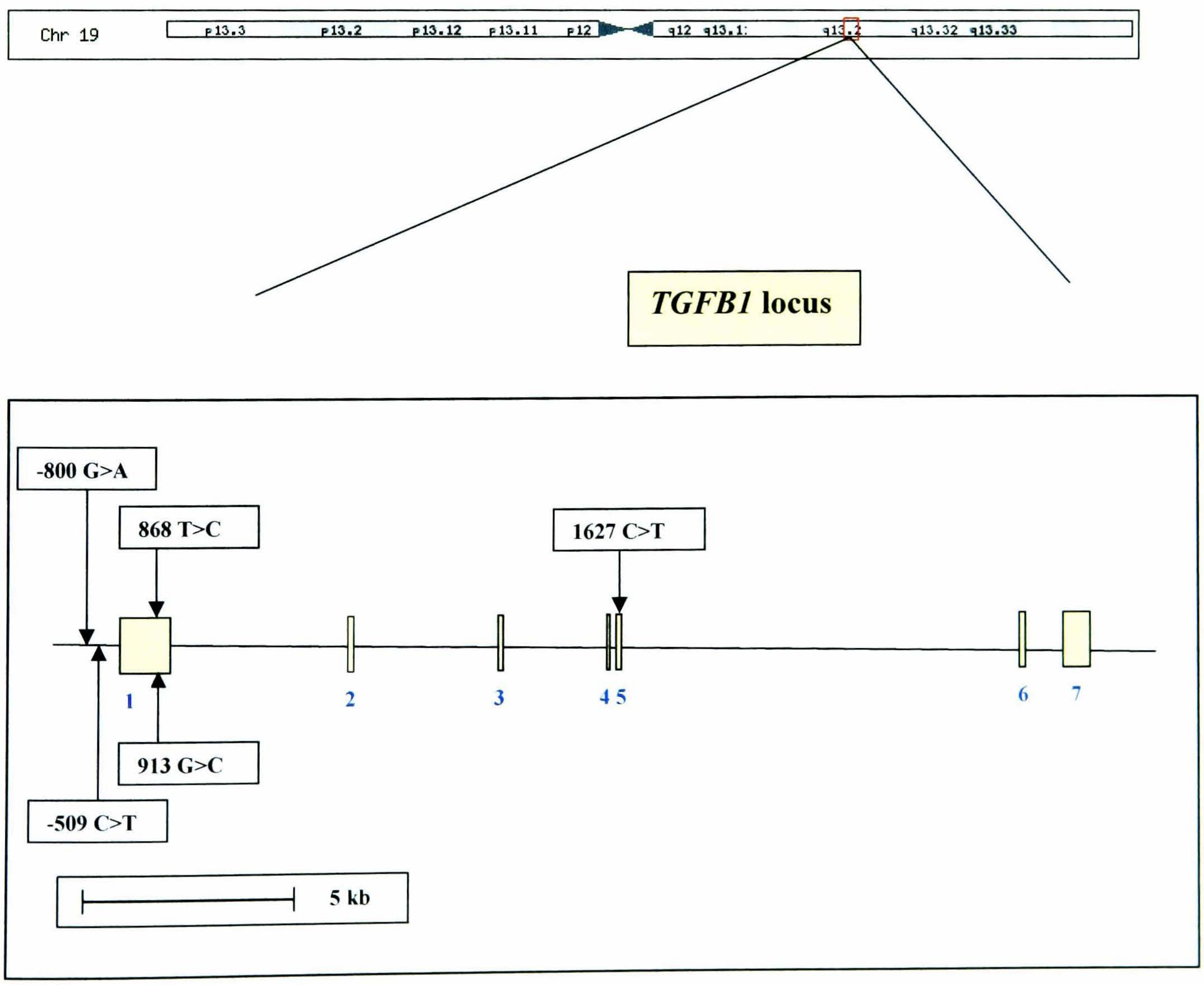
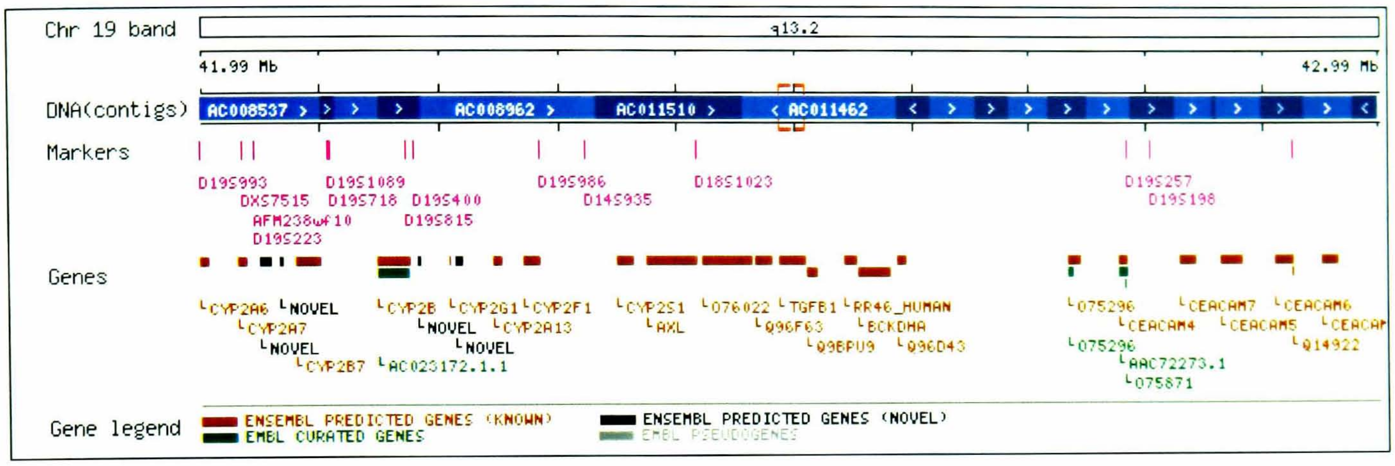


Figure 3.2 Genomic region around the TGFB1 gene.

The figure was obtained from the Ensembl webserver at www.ensembl.org.

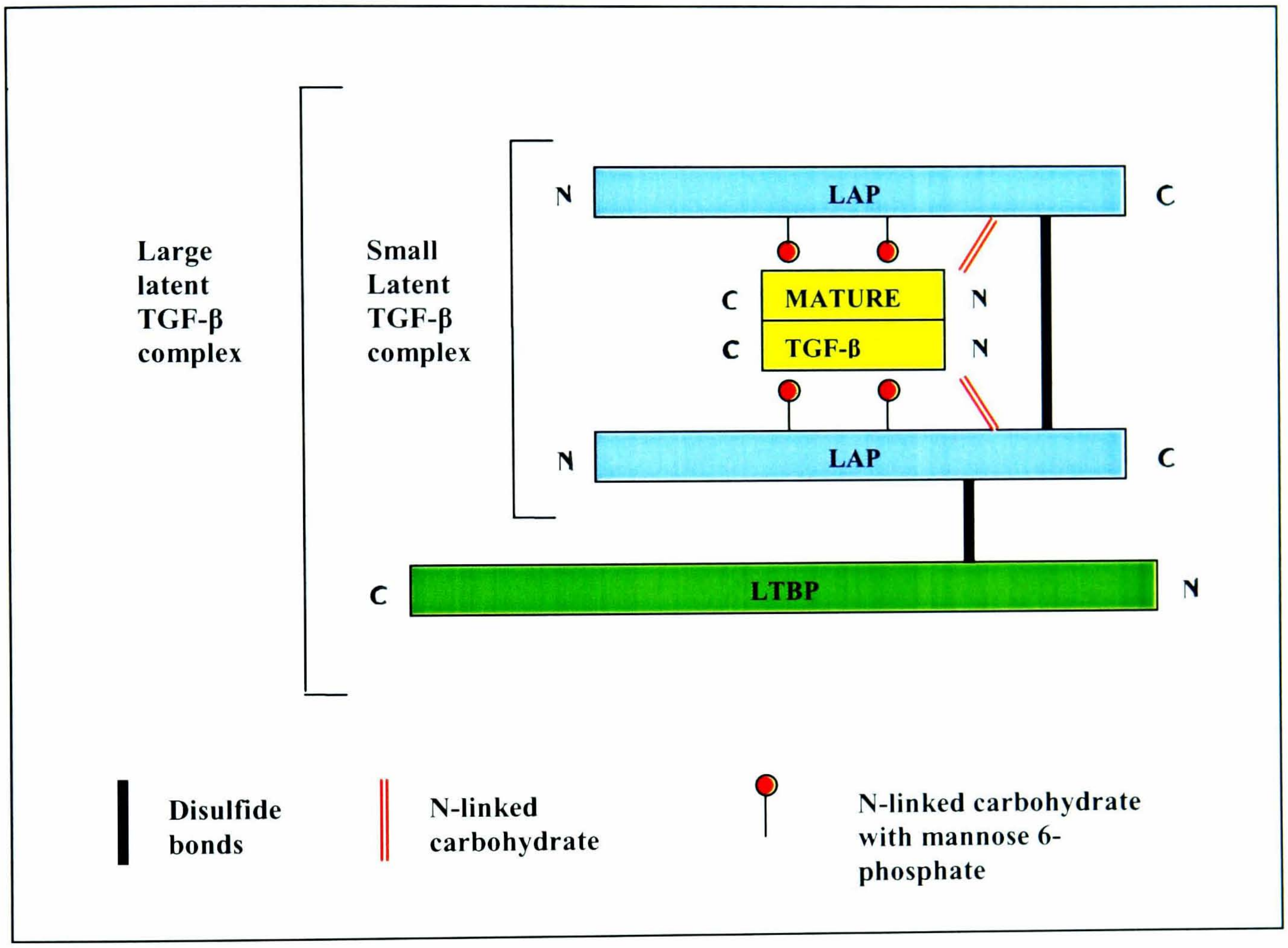


Three SNPs in the promoter region, one insertion/deletion in the 5' untranslated region, two SNPs in the signal peptide sequence, one SNP in the intron 4 and one SNP in the exon 5 have been identified previously (Cambien et al. 1996; Langdahl et al. 1997) (figure 3.1). Polymorphisms in the signal peptide sequence and in exon 5 are all nonsynonymous. Polymorphisms at positions 868 T>C and 913 G>C (reference sequence: ENST00000221930) in the signal protein sequence of the *TGFB1* gene change codon 10 (Leu10Pro) and codon 25 (Arg25Pro), respectively, and the polymorphism in exon 5 at position 1627 C>T changes the codon 263 (Thr263Ile).

TGF-β1 is made as a larger precursor of 390 amino acids, which contains an N-terminal signal peptide, a long prosegment called latency-associated polypeptide (LAP) and a 112-amino-acid C-terminal polypeptide that constitutes the mature TGF-β1 monomer (Derynck and Choy 1998). TGF-β1 is processed in the Golgi apparatus, where it is glycosylated, signal peptide is cleaved and dimers are formed. LAP remains associated non-covalently with the C-terminal fragment to form a latent molecule, which is unable to bind to receptors. Most cells secrete TGF-β1 in the form of large latent complexes containing additional proteins that associate with LAP. These additional proteins are called latent TGF-β-binding proteins (LTBPs). They are covalently linked to LAP by disulfide bonds (figure 3.3). Thus far four LTBP genes have been identified: *LTBP1, LTBP2, LTBP3* and *LTBP4* on 2p22-24, 14q22-q33, 11q12 and 19q13 respectively (Oklu and Hesketh 2000). Recently, it has been demonstrated that fibrillin-1, the protein that is defective in Marfan syndrome, has a key role in regulating TGF-β activation (Neptune et al. 2003). This finding supports the paradigm that structural matrix elements can also serve crucial regulatory roles in cytokine activation and signalling and perturbation of this process can contribute to disease pathogenesis. Mature TGF-β1 monomer is cleaved from the remaining precursor segment by a furin-like protease (Dubois et al. 1995). Biologically active forms of TGF-βs exist mainly as 25 kDa homodimers. Heterodimers between TGF-β1 and TGF-β2 have been isolated (Cheifetz et al. 1987), but their biologic relevance is unclear.

Figure 3.3 The structure of TGF-β latent complexes.

Small latent TGF-β consists of a mature TGF-β homodimer associated non-covalently with a LAP homodimer. The large latent TGF-β complex contains an additional protein, LTBP, which is covalently linked to LAP by disulfide binds. The figure is modified after (Oklu and Hesketh 2000).



The signal peptide is involved in the translocation of newly synthesised TGF-β1 across the endoplasmic reticulum and consists of a positively charged N-terminal region, a central hydrophobic core and a polar C-terminal region (Randall and Hardy 1989). Changes to the amino acid composition of the peptide could affect its polarity or structure and lead to different rates of protein export. Leu10Pro and Arg25Pro polymorphisms are located in the hydrophobic core of the signal sequence. Both leucine and proline at position 10 are apolar, but proline induces 90° kinks in the chain whereas leucine allows different conformations. A change from the big polar amino acid arginine to the small apolar proline at position 25 could potentially affect intracellular trafficking or export efficiency of the preproprotein (Verner and Schatz 1988). Several *TGFB1* polymorphisms have been demonstrated to be associated with altered serum concentrations of TGF-β1, suggesting a functional consequence. Furthermore, several common inflammatory and autoimmune diseases have been associated with these *TGFB1* polymorphisms (Pulleyn et al. 2001; Crilly et al. 2002; Sugiura et al. 2002).

Mutations in the *TGFB1* gene have also been associated with Camurati Engelmans disease, which is a rare sclerosing bone dysplasia inherited in an autosomal dominant manner (Janssens et al. 2000; Kinoshita et al. 2000). Four missense mutations Tyr81His, Arg218Cys, Arg218His, Cys225Arg and a triple insertion of three leucines in the signal peptide 35insGCTGCTGCT have been identified. These variants may impair the ability of the propeptide to maintain the mature TGF-β1 in its inactive state.

3.1.3 TGF-β receptors and signalling pathways

TGF-β exerts its effects by binding to specific high affinity transmembrane receptors. Three major receptors have been identified, designated type I, II and III TGF-β receptors, and they are ubiquitously distributed. Type I and type II receptors are transmembrane serine-threonine kinases that interact with one another and facilitate each other's signalling (Massague 1998). Both type I and type II receptors are required, but are not always sufficient for signal transduction to occur. Granulocytes, macrophages and lymphocytes possess functional TGF-β type I and II receptors but lack the type III receptor. Type III receptors, also called betaglycan and endoglin, do not appear to have an intrinsic signalling function but regulate TGF-β access to the signalling receptors. Endoglin is a cell surface molecule expressed at high levels in endothelial cells. Mutations in endoglin and ALK1 (a member of the family of type I TGF-β receptors) are responsible for an autosomal dominant vascular dysplasia, hereditary hemorrhagic telangiectasia type 1 and 2, respectively (McAllister et al. 1994; Johnson et al. 1996). Fibroblasts and other mesenchymal cells possess all three receptor types and are major targets of the biological activity of TGF-β in the regulation of extracellular matrix.

The downstream signalling of TGF-β is mediated through SMAD proteins. The term SMAD is derived from the founding members of this family, the *Drosophila* protein MAD (mothers against decapentaplegic) and the *Caenorhabditis elegans* protein SMA (small body size) (Massague 2000). SMADs are receptor-activated transcription factors that are phosphorylated in response to ligand binding. They are translocated to the nucleus following phosphorylation. In the nucleus SMADs interact with DNA-binding proteins to mediate transcriptional responses.

3.1.4 Function of TGF-β1

The major effects of TGF-β include production of extracellular matrix, stimulation of chemotaxis, and regulation of growth and differentiation of mesenchymal cells. TGF-β is chemoattractant for macrophages, neutrophils and lymphocytes. It can block the production of nitric oxide by macrophages, and both modulate the production of and antagonise the response of cells to inflammatory cytokines including TNF-α, IFNγ and various interleukins. TGF-β1 can initiate and terminate tissue repair. Its sustained production leads to the development of tissue fibrosis (Border and Noble 1994).

TGF-β1 is a potent regulator of the immune response and is generally immunosuppressive. It has profound suppressive effects on the growth of both T and B cells. It inhibits the differentiated function of mature B cells by suppressing expression of the immunoglobulins IgG and IgM, it has also been implicated in the IgA isotype switching. Mice with targeted disruption of the TGF-β1 gene display a phenotype that suggests the loss of a critical regulator of immune function. These mice develop a wasting syndrome accompanied by a multifocal, mixed inflammatory cell response and tissue necrosis, leading to organ failure and death (Shull et al. 1992).

TGF-β1 has a major effect on the immunological aspects of several diseases including autoimmune disease where expression of TGF-β1 has been implicated both in suppression of primary disease and in mediating effects of tolerance (Prud'homme and

Piccirillo 2000). TGF-β1 can profoundly inhibit the differentiation of precursor cells into dendritic cells promoting tolerance and reducing the inflammatory response (Hackstein et al. 2001). The autoimmune state may involve IFNγ induced expression of MHC class II molecules in a variety of cells (Bottazzo et al. 1983). TGF-β1 suppresses expression of class II MHC antigens induced by IFNγ (Czarniecki et al. 1988). Consequently, absence of TGF-β1 could lead to presentation of self antigens by inappropriate cells thereby eliciting an autoimmune response (Shull et al. 1992). Other diseases where TGF-β1 is likely to play a key role are carcinogenesis, parasitic infections and chronic inflammatory disease (Blobe et al. 2000).

TGF-β1 can exert a multitude of effects in all aspects of inflammation. A major determinant of the actions of this cytokine is the micro-environment in which TGF-β1 is found. Its effects are critically dependent on genetic background, cell type, state of cellular differentiation or activation, extracellular matrix, other growth factors and a plethora of additional coregulatory factors (Wahl 1999).

3.1.5 TGF-β1 in AS pathogenesis

There are several lines of evidence suggesting that TGF-β1 may play an important functional role in AS pathology (Archer 1995). TGF-β1 plays a crucial role in inflammatory processes, extracellular matrix synthesis, bone remodelling and fibrosis and may be important in the biological pathways related to the expression of AS. TGF-β1 is profibrotic and is possibly also responsible for late-stage fibrosis and ankylosis. TGF-β1 acts in the formation and repair of cartilage and bone (Noda and Camilliere 1989), which are the major targets of the immune response in AS. TGF-β1 promotes the IgA switching in B cells. Total serum IgA is elevated in AS patients (Veys and van Leare 1973), usually during the inflammatory phases of disease activity (Cowling et al. 1980).

Classically, T helper cells (Th) are functionally divided into two major subsets. Th1 lymphocytes produce IFNγ and IL-2 and mediate cellular immune responses, whereas Th2 lymphocytes, which produce IL-4 and IL-10, are implicated in humoral responses and allergy. AS is associated with impaired Th1 cytokine profile (Baeten et al. 2001). TGF-β1 can inhibit the production of, and response to, cytokines associated with both Th1 and Th2 cells (Letterio and Roberts 1998). TGF-β1 has multiple suppressive actions on T cells, B cells and macrophages. Increased TGF-β1 production inhibits autoimmune and chronic inflammatory diseases (Prud'homme and Piccirillo 2000). In animal models systemic delivery of TGF-β1 or intramuscular injection of a plasmid expressing TGF-β1 inhibits acute and chronic arthritis (Chen and Wahl 1999). The results of a study of polymorphisms in this functional and positional candidate gene on the susceptibility to, and severity of, AS in English and Finnish populations are presented in this chapter.

3.2 SUBJECTS AND METHODS

3.2.1 Families with AS

Four hundred and thirty-seven individuals from the 170 Finnish families were available for study. They included 12 affected sibling pairs, 11 parent-case affected pairs and 147 simplex families. Additionally, 1332 individuals from 212 parent-case trio families and 184 AS affected-sibling pair families recruited from the U.K. were used in the study. The affected-sibling pair families included altogether 421 patients with AS. The 212 parent-case trios comprised 31 parent-case affected pairs and 181 simplex families. Both the English and Finnish patients and families have been included in previous studies (Laval et al. 2001; Crane et al. 2002; Goedecke et al. 2003).

3.2.2 *HLA-B27*, *TGFB1* SNP and Microsatellite genotyping

Five polymorphisms located in the promoter and coding region of the *TGFB1* gene and three microsatellite markers flanking the *TGFB1* locus were genotyped in families with AS. One thousand and seven individuals from 212 English and 170 Finnish families with AS were genotyped for all the five SNPs. Seven hundred sixty-two individuals from 184 multiplex families with AS were typed for the two polymorphisms in the promoter region of the *TGFB1* gene and three microsatellite markers. The SNPs -800 G>A, -509 C>T, 868 T>C (commonly referred as 869 T>C), 913 G>C (commonly referred as 915 G>C) and 1627 C>T (reference sequence: ENST00000221930) were genotyped by PCR-SSP (see Appendix 3 for details of the primers and PCR conditions) (Lympany et al. 1998). Positive and negative controls were included in all the reactions. The same positive controls were used in both the Finnish and the English data set and all ambiguous samples were repeated. *HLA-B27* was detected by the method described by Bunce (Bunce et al. 1995). Families with *HLA-B27* positive proband were included in the analysis.

Three microsatellites spanning 6.4 cM around the *TGFB1* locus were genotyped: D19S421, D19S223 and D19S217 (see Appendix 3 for details of the primers and PCR conditions). The microsatellites were amplified under optimised PCR conditions, and the PCR products were separated and detected by capillary electrophoresis using an ABI Prism 3700 genotyper (Applied Biosystem, Warrington, UK). Products were then sized and manual inheritance checks performed using Genotyper version 1.1 (PE Biosystems) and Genescan version 2.1 (PE Biosystems).

3.2.3 Statistical analysis

Mendelian inheritance of markers was checked using the program GAS (version 2.0) [A. Young, unpublished]. The chi-square test with a 3 x 2 contingency table was used to examine whether the genotype frequencies were in HWE. P-values >0.05 were accepted as non-significant deviation from HWE. The program MAKEPED was used to produce a linkage format pedigree file that could be recognised by subsequent programs. The program DOWNFREQ was used to estimate the allele frequencies of the markers. MEGA2 (Mukhopadhyay et al. 1999) (version 2.3, http://watson.hgen.pitt.edu/mega2.html) was used to convert data files into the formats required for further analysis programs.

Association between *TGFB1* polymorphisms and the disease susceptibility was studied with the TDT using the program TRANSMIT (version 2.5) (Clayton 1999). P-values were obtained by statistical simulation using 1000 replicates, using the robust variance option to assess association independent of linkage. NPL analysis was performed using the program GENEHUNTER-PLUS for qualitative data (Kong and Cox 1997) and MERLIN-REGRESS for quantitative data (Abecasis et al. 2002). Association was evaluated in a subset of families that showed positive NPL score. Association between *TGFB1* polymorphisms and the age of symptom onset, BASDAI and BASFI was assessed using the QTDT (Abecasis et al. 2000). Population stratification and total evidence of association were assessed using QTDT.

TGFB1 haplotypes were constructed using the PHASE program (Stephens et al. 2001). Mendelian inheritance of haplotypes was manually checked and only haplotypes which were constructed with a greater than 90% certainty were accepted for further analysis. These haplotypes were used as an input for QTDT in order to assess the association between *TGFB1* haplotypes and the age of symptom onset, BASDAI and BASFI. BASFI and BASDAI were correlated with disease duration and gender in this dataset and were treated as covariates in the analysis. The association between the polymorphisms and the presence of AAU, peripheral arthritis or IBD was studied using the chi-square test. For the interethnic comparison, the haplotypes were constructed using the program TRANSMIT (Clayton 1999).

The pair-wise LD between the SNPs was calculated using the Lewontin's standardised disequilibrium coefficient D' (Lewontin 1988), calculated using the program HAPLOXT (Abecasis and Cookson 2000). The founder haplotypes estimated using the program TRANSMIT were used as the input (Clayton 1999). The program Entropy was used to identify the subset of SNPs that best approximates the haplotypic diversity in the population [www.well.ox.ac.uk/~rmott/SNPS]. The power of the study to detect association with susceptibility to AS was determined using TDT Power Calculator (Chen and Deng 2001). To determine whether the promoter polymorphisms -800 and -509 lie in a regulatory region, the PipMaker program (http://bio.cse.psu.edu/pipmaker/) was used to compare the mouse and human *TGFB1* sequences (Schwartz et al. 2000). Sequence homology could indicate evolutionary conserved regulatory regions.

3.3 RESULTS

3.3.1 Clinical description of the families

The clinical characteristics of the AS patients are described in the table 3.1.

Table 3.1 Clinical description of the AS patients.

This table includes only those patients from whom all the clinical data was available. In the English data set the clinical data was available from 86% of the 664 patients used in this study. In the Finnish data set the clinical data was available from 99% of the 193 patients used in this study.

	ENGLISH AS	FINNISH AS
BASFI ± SD	3.6 ± 2.7	3.7 ± 2.2
BASDAI ± SD	3.9 ± 2.1	5.2 ± 1.8
AGE OF SYMPTOM ONSET ± SD	22.4 ± 7.6	26 ± 8.0
DISEASE DURATION ± SD	22.0 ± 13	20 ± 10
MALES (%) : FEMALES (%)	356 (63) : 212 (37)	127 (66) : 64 (34)
AAU (%)	214 (38)	91 (48)
IBD (%)	34 (6)	12 (6)

3.3.2 Disease susceptibility results

Transmitted and untransmitted marker and SNP haplotype frequencies estimated using the program TRANSMIT are given in tables 3.2 - 3.4. By within-family means, a marginal association was noted between the rare *TGFB1* 1627 T allele and AS in a Finnish population (4.2% vs. 5.3%, p=0.04) and in the combined data set (2.8% vs 3.7%, p=0.03), but not in the English data set alone (1.7% vs. 1.9%, p=0.34). A marginal association was also noted between the promoter -800-509 haplotype 2/1 and AS in the English population (p=0.05), but this was neither significant in the Finnish population nor in the combined data set. No association was noted between any other SNP or marker haplotype and AS in either population.

Table 3.2 ***TGFB1*** **marker untransmitted and transmitted founder allele frequencies and TDT p-value.**

Marker	Ethnic Group	Allele		Untransmitted founder Alleles N	Untransmitted founder Alleles %	Transmitted founder Alleles N	Transmitted founder Alleles %	P-value
-800	Finnish	1	G	374	94	373	94	0.68
		2	A	24	6	25	6	
	English	1	G	1115	91	1128	92	0.06
		2	A	109	9	96	8	
-509	Finnish	1	C	297	75	294	74	0.51
		2	T	101	25	104	26	
	English	1	C	905	74	901	74	0.65
		2	T	319	26	323	26	
868	Finnish	1	T	294	74	292	74	0.67
		2	C	102	26	104	26	
	English	1	T	267	65	260	63	0.29
		2	C	143	35	150	37	
913	Finnish	1	G	378	95	381	96	0.26
		2	C	18	5	15	4	
	English	1	G	375	91	377	92	0.7
		2	C	37	9	35	8	
1627	Finnish	1	C	378	96	373	95	0.04
		2	T	16	4	21	5	
	English	1	C	406	98	404	98	0.34
		2	T	8	2	10	2	

Table 3.3 ***TGFB1*** **SNP (-800, -509, 868, 913, 1627) untransmitted and transmitted founder haplotype frequencies in the Finnish and English population.**

(Only haplotypes with the frequency > 1% are shown.)

1 = wild type allele

2 = rare allele

ENGLISH

	Untransmitted founder Haplotypes		Transmitted founder Haplotypes		
Haplotype	N=414	%	N=414	%	p-value
11111	221	53	227	55	0.39
12211	99	24	104	25	0.41
21111	35	8	28	7	0.7
11221	28	7	32	8	0.35
12212	8	2	9	2	0.44
11211	9	2	6	1	0.23
11121	7	2	4	1	0.21
12111	4	1	2	0.5	0.27

FINNISH

	Untransmitted founder Haplotypes		Transmitted founder Haplotypes		
Haplotype	N=388	%	N=388	%	p-value
11111	243	63	246	63	0.62
12211	72	19	74	19	0.63
21111	22	6	24	6	0.65
11221	12	3	9	2	0.29
12212	12	3	15	4	0.26
12111	12	3	9	2	0.24
12112	4	1	2	0.5	0.2

Table 3.4 *TGFB1* SNP promoter (-800, -509) untransmitted and transmitted founder haplotype frequencies.

ENGLISH

	Untransmitted founder Haplotypes		Transmitted founder Haplotypes		
Haplotype	N=1230	%	N=1230	%	p-value
11	799	65	809	66	0.34
12	322	26	325	26	0.73
21	109	9	96	8	0.05

FINNISH

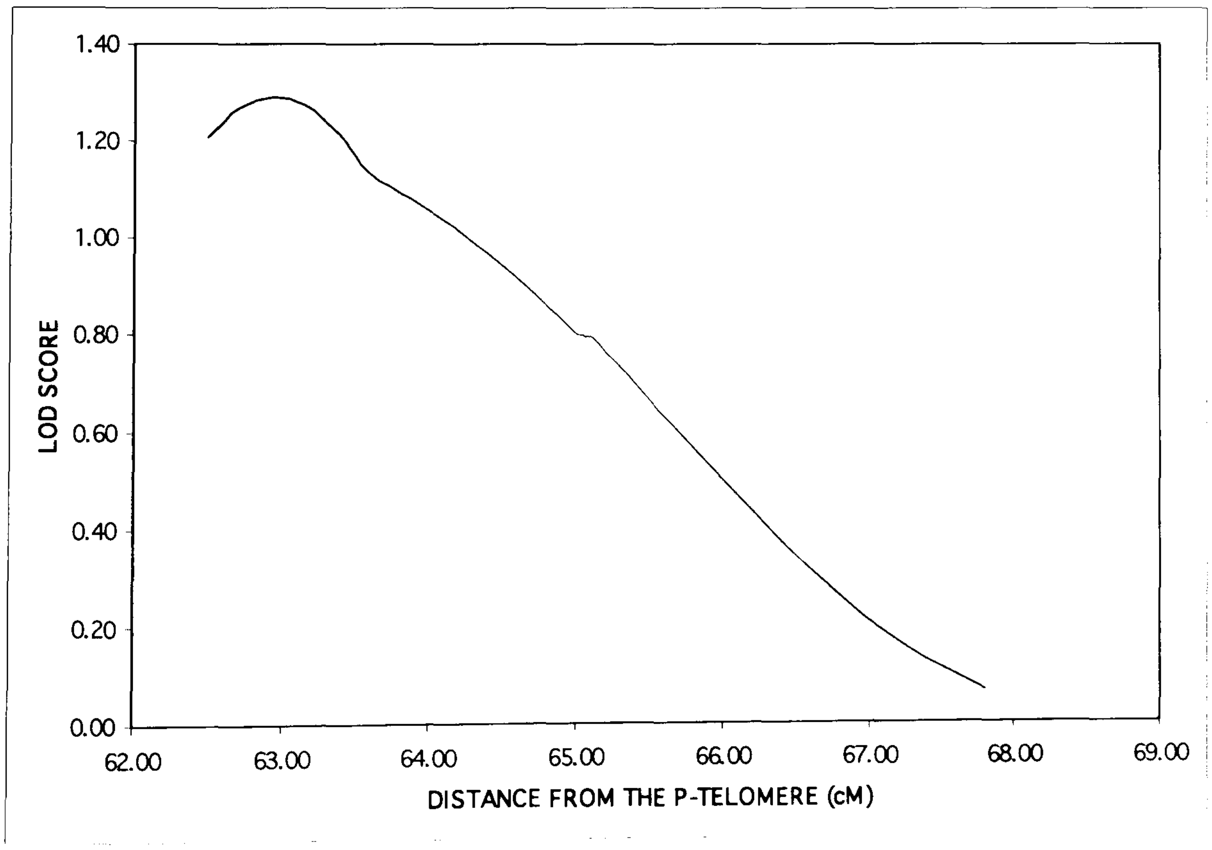
	Untransmitted founder Haplotypes		Transmitted founder Haplotypes		
Haplotype	N=372	%	N=372	%	p-value
11	257	69	253	68	0.45
12	94	25	97	26	0.58
21	20	5	22	6	0.53

3.3.3 NPL analysis and evaluation of association among the families showing linkage to the *TGFB1* locus

The multipoint NPL analysis using GENEHUNTER-PLUS showed a weak evidence of linkage centromeric from the *TGFB1* locus (figure 3.4). The highest LOD score was obtained 2 cM centromeric from *TGFB1* gene between the markers D19S421 and D19S223, the peak LOD score being 1.3 (p=0.02) at distance 63 cM from the p-telomere. Ninety-six families showed a positive NPL score, but no association was detected between AS and *TGFB1* -800 or -509 polymorphisms among these families. *TGFB1* 1627 C>T polymorphism, which showed a weak association, was not genotyped in the subgroup of families with positive NPL score due to the LD between this SNP and the promoter polymorphisms.

Figure 3.4 Multipoint non-parametric linkage results using the program GENEHUNTER-PLUS.

TGFB1 gene lies 65 cM from the telomere of the p arm of chromosome 19.



3.3.4 Disease severity results

Using the total evidence of association option within the QTDT program, association was noted between age of symptom onset and *TGFB1* polymorphisms in different populations. However this was not consistently observed in the three family collections. In the English parent-case trio families, association with age of onset was noted with *TGFB1*-509 (allele 2 associated with increased age of onset by 2.1 years, p=0.007) and 1627 (allele 2 associated with increased age of onset by 4.2 years, p=0.05). In this data set association was also noted between age of symptom onset and haplotypes of the promoter region variants -800/-509 (haplotype 1/1 associated with decreased age of onset by 1.8 years, p=0.02; haplotype 1/2 associated with increased age of onset by 2.1 years, p=0.007), and with haplotypes of the coding region variants 868/913/1627 (haplotype 1/1/1 associated with decreased age of onset by 1.7 years, p=0.01; haplotype 2/1/1 associated with increased age of onset by 1.8 years, p=0.03; haplotype 2/1/2 associated with increased age of onset by 4.9 years, p=0.03). The promoter region associations were not replicated in the English affected sibling pair families, and when all English families were analysed together, no significant association was noted for *TGFB1* -800, -509 or their haplotypes. Pooling English and Finnish families genotyped for the coding region polymorphisms, weak association was noted between *TGFB1* 913 (allele 1 associated with increased age of onset by 2.5 years, p=0.01) and 1627 (allele 2 associated with increased age of onset by 3.2 years, p=0.02) variants. In the combined data set association was also noted between age of onset and haplotypes of coding region polymorphisms 868/913/1627 (haplotype 2/2/1 associated with decreased age of onset by 2.3 years, p=0.05; haplotype 2/1/2 associated with increased age of onset by 4.2 years, p=0.006). No significant association was observed with age of symptom onset in the Finnish families, although the dataset was too small to provide adequate power to exclude a significant association, particularly of haplotypes.

No effect of population stratification on these associations was noted in the dataset. No associations were noted between BASDAI or BASFI and any of the polymorphisms. No linkage was found between quantitative traits and the SNPs. No significant association between the polymorphisms and the presence of AAU, peripheral arthritis or IBD was noted in this study.

3.3.5 Power of the study

The power of this study was calculated for dominant, codominant and recessive inheritance models (table 3.5), assuming an AS prevalence of 0.1% and type I error of 0.05. For dominant and codominant models, the study had 80% power to detect association of loci with genotype relative risks of 1.4-3.1 depending on the SNP concerned. For recessive models, the power was poor for the *TGFB1* 913 and 1627 polymorphisms (80% power for genotype relative risks of 10.2-26), but moderate for the other SNPs (80% power for genotype relative risks of 1.9-4.6).

Table 3.5 Power of the study.

This study had a 0.80 power to detect an association with the following GRRs, the type I error of 0.05 and AS prevalence of 0.1%.

SNP	DISEASE	GRR
-800	dominant	1.5
	codominant	1.8
	recessive	4.6
-509	dominant	1.4
	codominant	1.5
	recessive	1.9
868	dominant	1.6
	codominant	1.8
	recessive	2.2
913	dominant	1.8
	codominant	2.5
	recessive	10.2
1627	dominant	2.1
	codominant	3.1
	recessive	26.0

3.3.6 LD and haplotype studies

The founder allele frequencies between the Finnish and English data sets and the control individuals in two previously published British studies were compared (Awad et al. 1998; Syrris et al. 1998) (table 3.6). No significant differences in allele frequencies were noted between the English data set and the historic studies. The allele frequencies of *TGFB1* -800, 868 and 913 polymorphisms showed significant differences between the Finnish data set and the current English data set (p-values 3×10^{-3}, 10^{-8} and 5×10^{-5}, respectively). Significant differences were also noted between the Finnish data set and two historic British studies (table 3.6).

Of the 32 possible SNP-haplotypes, two main haplotypes accounted for 79% of all haplotypes (table 3.7), consistent with previous reports (Syrris et al. 1998; Tiret et al. 2002). Eight haplotypes with a frequency of more than 1% were identified. Significant differences were noted in haplotype frequencies between the Finnish and English population (chi-squared test of homogeneity = 107 [8 degrees of freedom], p-value $<10^{-10}$).

Table 3.8 shows the D' and the corresponding p-values between the SNPs in the two populations. By inspection, there are significant differences in the LD patterns between the English and Finnish populations. Figure 3.5 demonstrates graphical overview of LD between the SNPs within and in the promoter region of *TGFB1* gene created using the program GOLD.

Entropy values for different combinations of SNPs are given in figure 3.6. The two promoter SNPs and the *TGFB1* 868 T>C polymorphism account for 85% and 87% of the total diversity observed for all five SNPs in the Finnish and English populations respectively.

Table 3.6 Allele frequencies in this study and in two previously published British studies (above) and comparison between the studies (below).

Allele frequencies

Author			This study	This study	Syrris *et al.*	Awad *et al.*
Population			UK	Finland	UK	UK
			alleles (%)	alleles (%)	alleles (%)	alleles (%)
-800	1	G	2388 (90)	810 (94)	450 (92)	185 (89)
	2	A	256 (10)	54 (6)	38 (8)	23 (11)
-509	1	C	1972 (75)	652 (76)	345 (70)	158 (76)
	2	T	672 (25)	210 (24)	143 (30)	50 (24)
868	1	T	727 (65)	660 (77)	313 (64)	139 (65)
	2	C	395 (35)	200 (23)	175 (36)	75 (35)
913	1	G	1022 (92)	829 (96)	458 (93)	193 (90)
	2	C	94 (8)	33 (4)	30 (7)	21 (10)
1627	1	C	1114 (98)	839 (97)	481 (98)	ND
	2	T	24 (2)	25 (3)	7 (2)	ND

ND = not determined

Comparison between the studies and corresponding p-values

Two studies compared	This study - UK data	This study - UK data	This study - Finnish data	This study - UK data	This study - Finnish data
	This study - Finnish data	Syrris *et al.*	Syrris *et al.*	Awad *et al.*	Awad *et al.*
	p-value	p-value	p-value	p-value	p-value
-800	0.0026	0.22	0.33	0.60	0.024
-509	0.57	0.08	0.05	0.72	0.99
868	1.2×10^{-8}	0.84	9.6×10^{-7}	0.97	5.6×10^{-4}
913	5.3×10^{-5}	0.14	0.07	0.60	6.4×10^{-4}
1627	0.33	0.48	0.13	ND	ND

Table 3.7 Localisation of SNPs and identification of haplotypes of the *TGFB1* gene.

Haplotypes with a frequency greater than 1% are shown. [There were 2598 haplotypes in the English cohort and 858 haplotypes in the Finnish cohort].

Nucleotide:	-800	-509	868	913	1627	Frequency	Frequency
Alleles:	G>A	C>T	T>C	G>C	C>T	N (%)	N (%)
Haplotype						English	Finnish
1	G	C	T	G	C	1351 (52)	532 (62)
2	G	T	C	G	C	650 (25)	154 (18)
3	A	C	T	G	C	208 (8)	51 (6)
4	G	C	C	C	C	182 (7)	26 (3)
5	G	C	C	G	C	52 (2)	5 (0.6)
6	G	T	T	G	C	26 (1)	34 (4)
7	G	T	C	G	T	52 (2)	26 (3)
8	G	C	T	C	C	26 (1)	9 (1)
9	G	T	T	G	T	0	9 (1)
Location:	5'	5'	codon 10	codon 25	codon 263		
Amino acid change			Leu/Pro	Arg/Pro	Thr/Pro		

Table 3.8 LD values between the SNPs in the Finnish and English population.

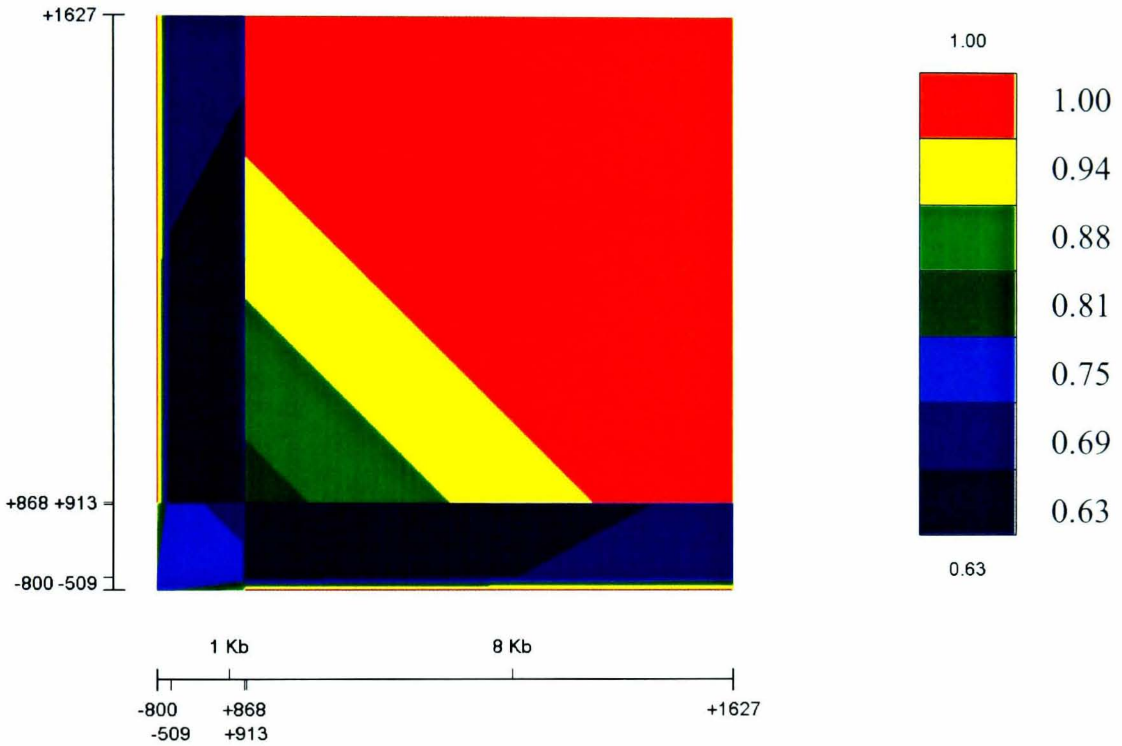
The table shows Lewontin's standardised disequilibrium coefficient D' (p-value). All calculations are based on 1 degree of freedom.

Polymorphism	Ethnic Group	-509 C>T	868 T>C	913 G>C	1627 C>T
-800 G>A	Finnish	0.88 (0.002)	0.88 (0.002)	0.41 (0.6)	0.39 (0.6)
	English	0.79 (0)	0.87 (0)	1 (0.007)	1 (0.2)
-509 C>T	Finnish		0.64 (0)	0.74 (0.02)	0.76 (0)
	English		0.77 (0)	0.63 (0.0008)	0.73 (0)
868 T>C	Finnish			0.58 (0)	0.46 (0)
	English			0.64 (0)	0.69 (0.0002)
913 G>C	Finnish				0.3 (0.7)
	English				1 (0.2)

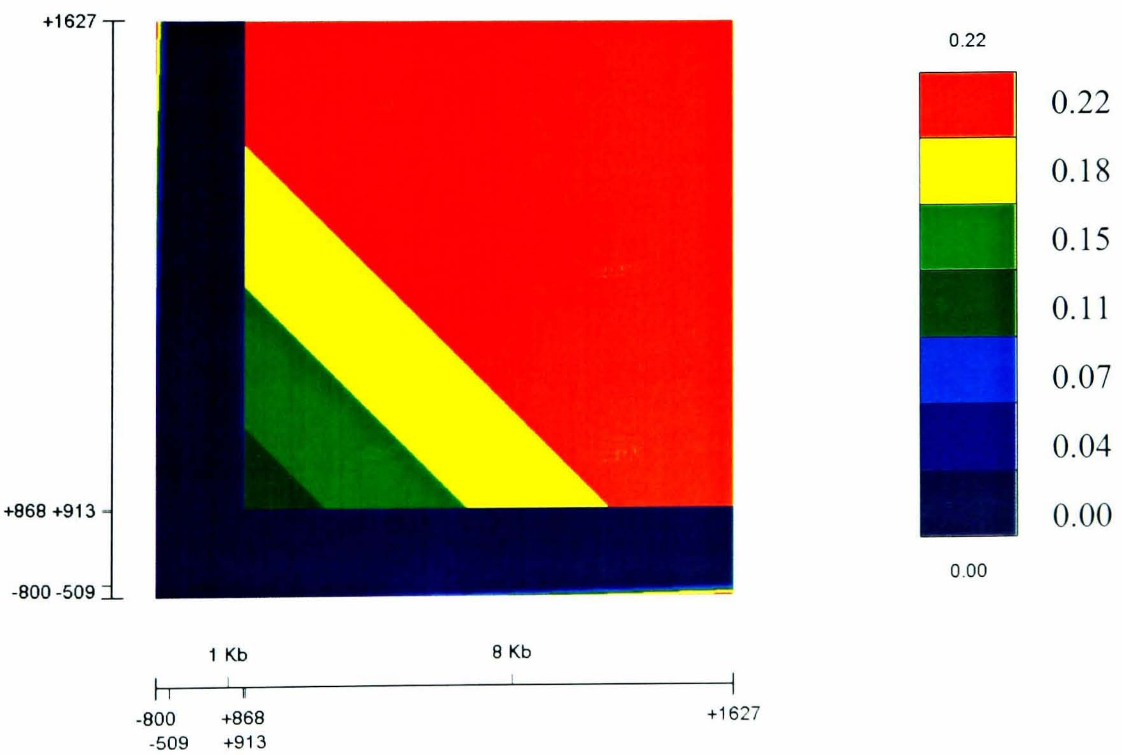
Figure 3.5 Graphical representation of LD between the SNPs in the *TGFB1* gene region (above) and the corresponding p-values (below).

The LD is described by Lewontin's standardized disequilibrium coefficient D', which measures the pair-wise LD between the SNPs. The figure presents the graphical overview of the patterns of disequilibrium in the *TGFB1* region and their relationship to the underlying physical map. Founder haplotype estimates were used as an input. The horizontal and vertical axes are scaled according to physical distances between markers. For each marker pair m_i and m_j, the pair-wise disequilibrium statistics are colour coded (bright red and dark blue are opposite ends of the scale) and plotted at position $[x, y] = [pos(m_i), pos(m_j)]$, where pos denotes the marker locations on the physical map.

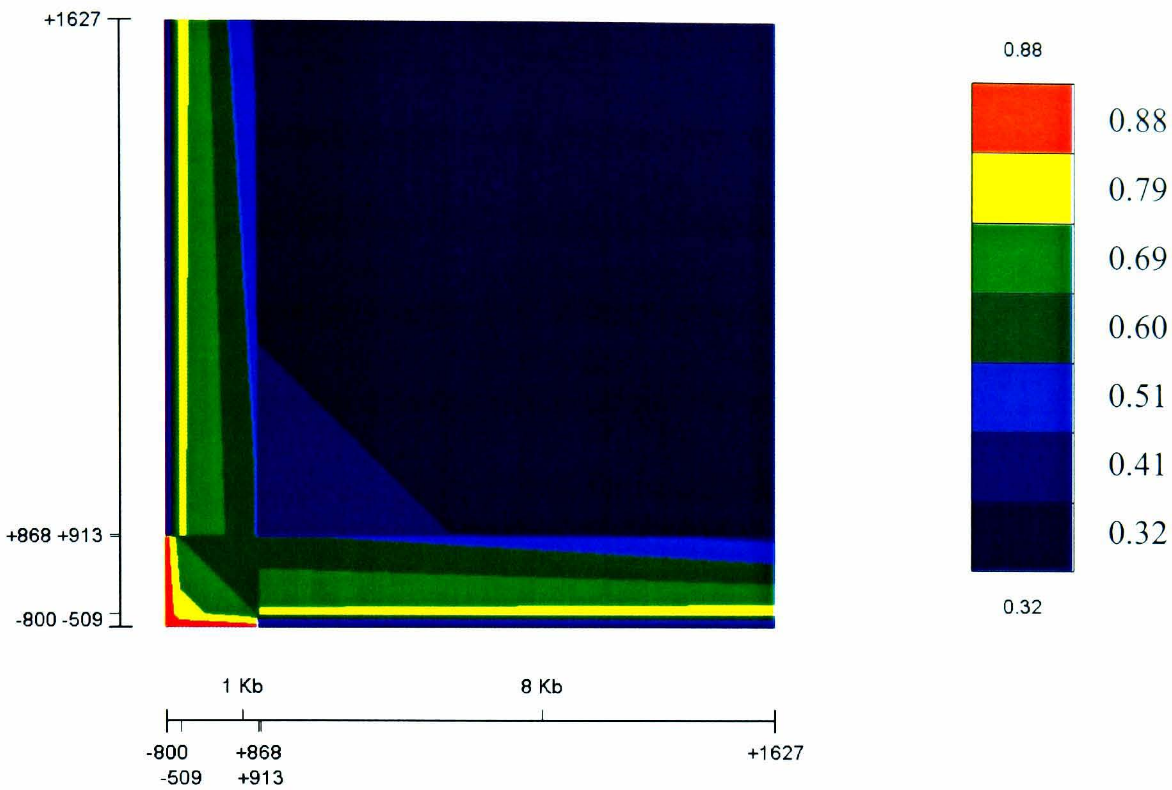
ENGLISH POPULATION – D'



corresponding p-values



FINNISH POPULATION – D'



corresponding p-values

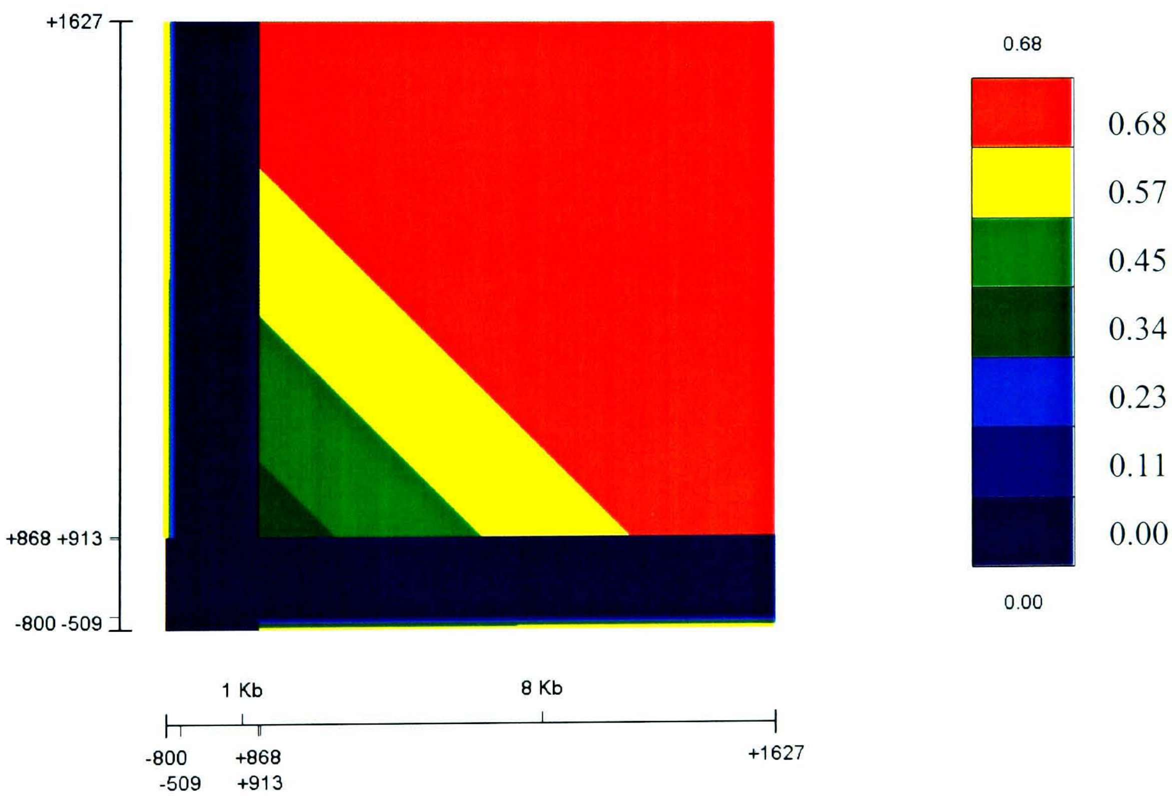


Figure 3.6 Values of entropy in the Finnish and English datasets.

The values are given for the following SNPs:

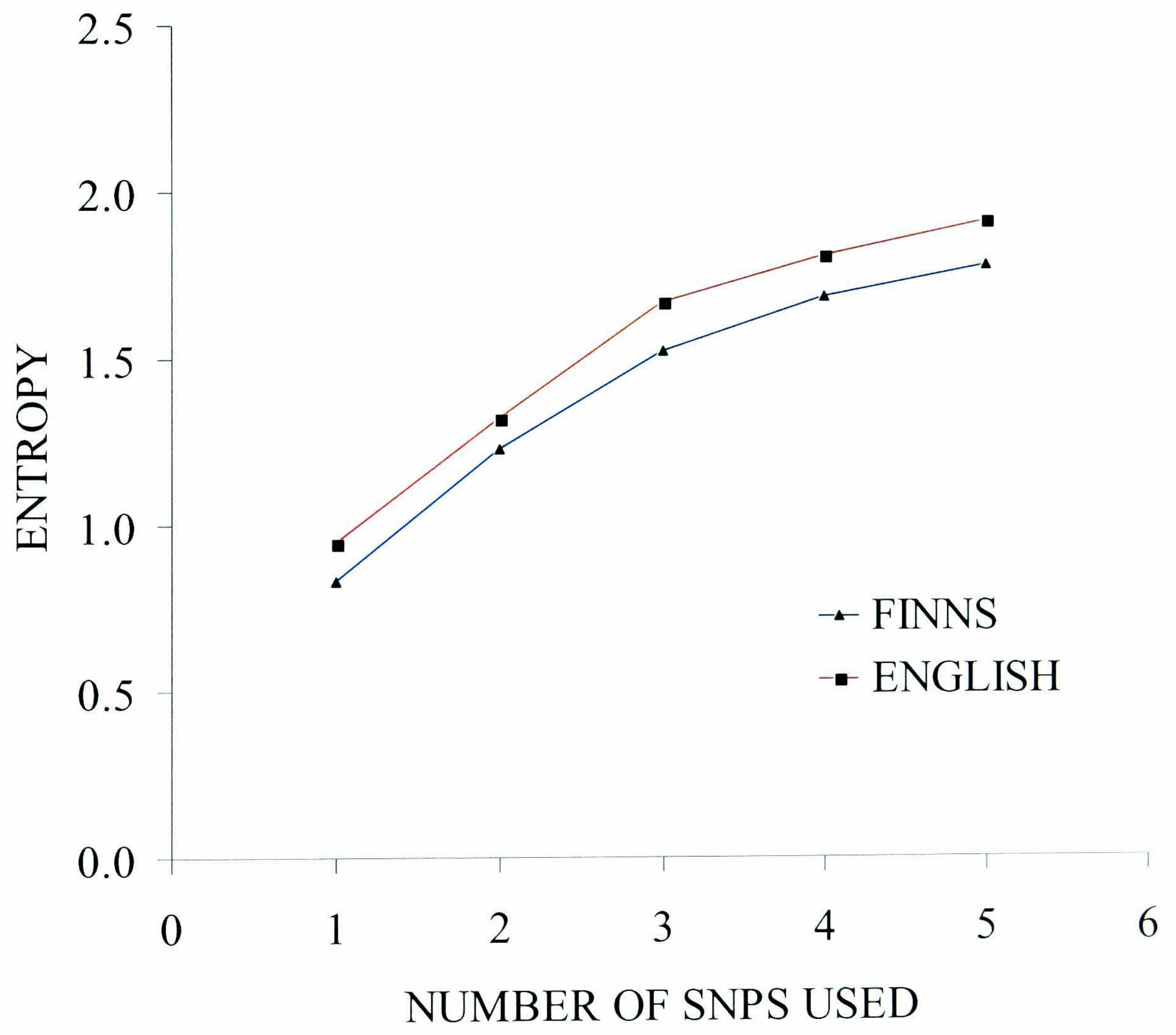
1 = 868 T>C

2 = 868 T>C and -509 C>T

3 = 868 T>C,-509 C>T and -800 G>A

4 = 868 T>C, -509 C>T, -800 G>A and 913 G>C

5 = 868 T>C, -509 C>T, -800 G>A, 913 G>C and 1627 C>T



3.3.7 Comparison of the human *TGFB1* sequence with mouse *Tgfb1*

Sequence alignment between human and mouse is demonstrated in figure 3.7 (reference number of the Ensembl mouse *Tgfb1* gene is ENSMUSG00000002603). The longest 20 homologous segments are represented in table 3.9. All *TGFB1* exons are conserved, and segments of sequence homology were noted in the promoter region and in the first intron. CpG islands were noted in the 5' region of the gene.

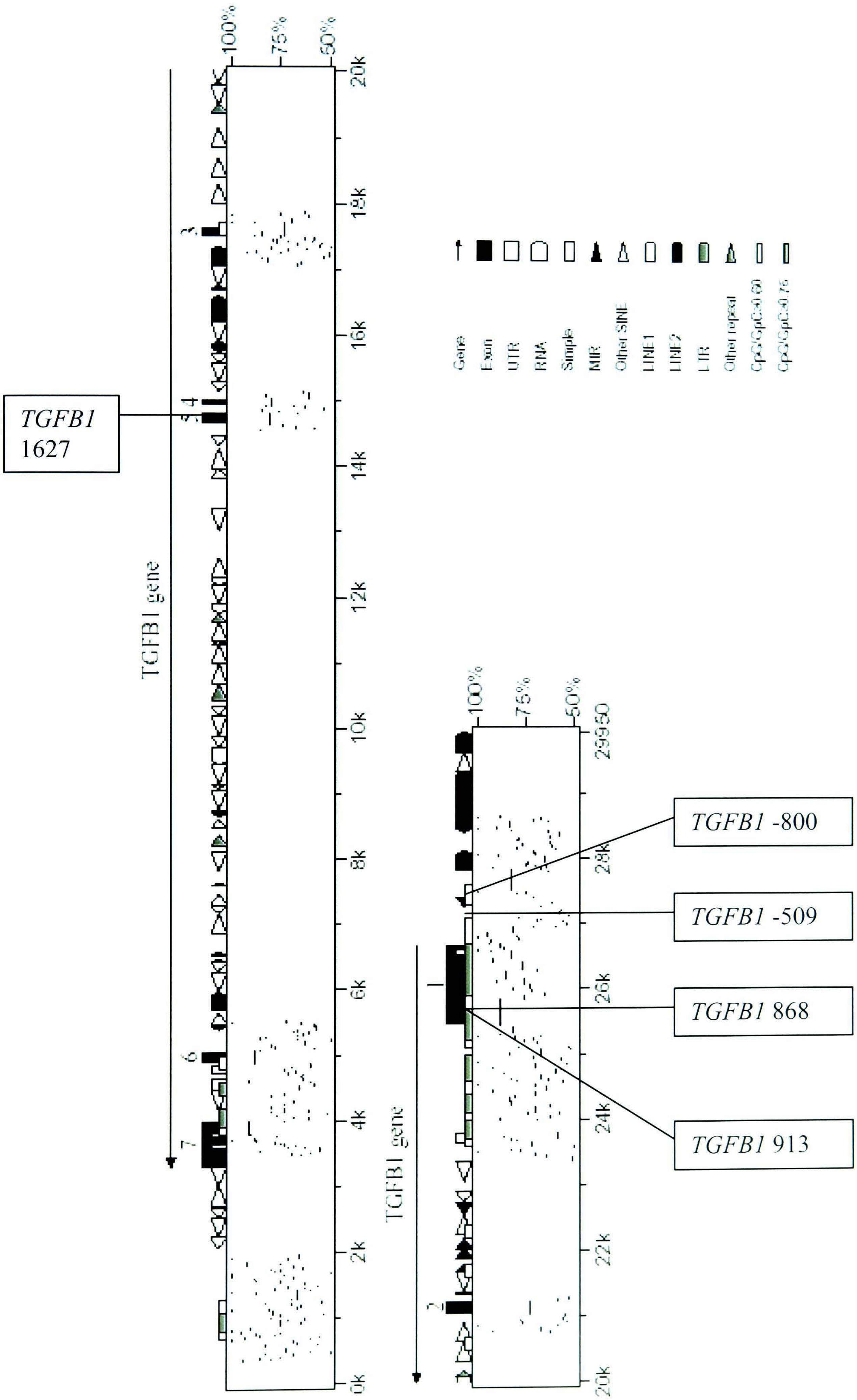
Table 3.9 The longest 20 homologous portions noted by the PipMaker program.

Segments	% of homology	Length in bp	Position	Position in bp*
1	88%	423	Exon 1	800-1222
2	83%	284	Promoter region	-862-(-1145)
3	89%	193	Exon 6	21583-21775
4	91%	190	Exon 7	22663-22852
5	73%	190	Exon 3	8951-9140
6	73%	189	Exon 2	5426-5614
7	81%	168	Exon 5	11850-12017
8	85%	110	Exon 1	396-505
9	77%	106	Exon 1	252-357
10	83%	100	Exon 1	78-177
11	82%	98	Intron 1	1224-1321
12	80%	92	Exon 4	11627-11718
13	84%	92	Promoter region	-1917-(-3020)
14	66%	80	Promoter region	-1303-(-1382)
15	57%	79	Intron 5	21199-21277
16	82%	79	Promoter region	-209-(-287)
17	72%	74	Intron 1	2991-3064
18	70%	69	Intron 1	2453-2521
19	84%	69	Intron 1	2311-2379
20	67%	66	Intron 1	1963-2028

*Transcription start site is denoted as 0.

Figure 3.7 Percent identity plot (following page) of the *TGFB1* genomic sequences between human and mouse by using the PipMaker program.

The human genomic sequences are shown on the x-axis and the percentage sequence identities (50-100%) are shown on the y-axis. Seven *TGFB1* exons are illustrated at the top of the sequences by solid black boxes and numbered 1-7. The arrow above the pip indicates the orientation of the *TGFB1* gene. The location of the polymorphisms genotyped in this study is indicated. The repeats of human sequence are depicted as follows: (black triangles) MIRs; (light gray triangles) SINEs other than MIRs; (light gray boxes) LINE1s; (black boxes) LINE2; (dark gray boxes) LTRs; (dark gray triangles) other repeats; (white boxes) simple repeats. Short dark gray boxes are CpG islands where the ratio of CpG/GpC exceeds 0.75 and short white boxes are CpG islands where the ratio of CpG/GpC is between 0.60 and 0.75.


TGFB1 gene
TGFB1 1627
7
6
5
4
3
2
1
100%
75%
50%
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
22k
24k
26k
28k
29950
TGFB1 913
TGFB1 868
TGFB1 -509
TGFB1 -800
Gene
Exon
UTR
RNA
Simple
MIR
Other SINE
LINE1
LINE2
LTR
Other repeat
CpG/GpC>0.60
CpG/GpC>0.75

3.4 DISCUSSION

TGFB1 is an attractive positional and functional candidate gene in AS due to its position under the chromosome 19 linkage peak and involvement in inflammatory processes, tissue fibrosis and bone remodelling. Suggestive linkage of AS to chromosome 19 was identified in a whole genome screen (Laval et al. 2001), the maximum overall LOD score detected being 3.5. A recent North American Spondylitis Consortium genome screen in AS also detected linkage on chromosome 19 giving further evidence that genes on chromosome 19 are involved in AS (Reveille 2002). A genome-wide screen in IBD has identified strong linkage to chromosome 19 (Rioux et al. 2000), which may be a relevant finding in AS pathology as subclinical ileitis is found in 49% of patients with AS (Leirisalo-Repo et al. 1994) and these diseases may share common underlying predisposing genes. This study noted a weak association between the rare *TGFB1* 1627 T allele and susceptibility to, and age of symptom onset of, AS. A significant association between the *TGFB1* -800-509 haplotype 2/1 and disease susceptibility was noted in the English population, but this was not significant in the Finnish population. However the lack of association of *TGFB1* promoter polymorphisms, which are in LD with the *TGFB1* 1627 alleles, with AS in families with positive NPL scores at the locus indicates that these polymorphisms do not explain the observed linkage of disease susceptibility to chromosome 19. It should be noted that the p-values reported here are not corrected for the multiple comparisons involved, as the number of independent comparisons would be difficult to determine. However the positive p-values reported would not be significant if such correction was employed, and these findings will require further confirmation.

Polymorphisms in *TGFB1* have been implicated in a variety of musculoskeletal diseases including rheumatoid arthritis (Sugiura et al. 2002), systemic sclerosis (Crilly et al. 2002) and osteoporosis (Langdahl et al. 1997; Yamada et al. 1998). An association between *TGFB1* 868 CC genotype and both the ossification of the posterior longitudinal ligament (OPLL) in the cervical spine and spinal osteophytosis has been reported in Japanese patients (Yamada et al. 2000; Kamiya et al. 2001). OPLL demonstrates ectopic ossification of the spinal ligaments very similar to that of AS (Mori et al. 2003). The prevalence of OPLL in AS has been reported to be 15% and seems to be associated with variables identifying more severe axial disease (Ramos-Remus et al. 1998). No association between IBD or Crohn's disease and *TGFB1* polymorphisms was found in previous studies (Garcia-Gonzalez et al. 2000; Schulte et al. 2001). A recent study from Scotland reported an association between *TGFB1* 913 polymorphism and AS and demonstrated an association of the CC genotype with higher serum concentrations of TGF-β1 (McGarry et al. 2001a). A weak association between age of symptom onset and this polymorphism was noted in both the combined and in the Finnish dataset. These findings show that this *TGFB1* polymorphism may play a role in AS, though a very weak one. This study failed to replicate the association with *TGFB1* 913 polymorphism and AS susceptibility. Marked differences in LD patterns were noted between the two populations. It is possible that the difference between the observations in the English and Finnish populations reported here and the difference between these results and those previously reported in a Scottish population may reflect differential association with a further truly associated polymorphism.

In searching for disease genes for complex traits, it has been suggested that substantial advantage can be obtained by gaining access to multiple populations with divergent demographic histories (Varilo et al. 2003). This study of *TGFB1* polymorphisms was in a Finnish and English population, both of Caucasian origin and with well-characterized population histories. Marked differences in allele and haplotype frequencies, disease associations and LD pattern were noted between these two populations. The results suggest that these populations have distinct population histories and that different *TGFB1* haplotypes have been traded differently by genetic drift. Previous studies have demonstrated that disequilibrium levels are not substantially greater in the general Finnish population than in the UK population (Eaves et al. 2000; Taillon-Miller et al. 2000). However, these two previous studies were limited to two chromosomal regions. LD varies considerably across the genome and further studies are required to confirm this observed pattern genome-wide. This is the first study to investigate the disequilibrium between a Finnish and English population at the *TGFB1* locus. Additionally, small LD differences, such as those observed in this study, especially within genes, may be relevant to the location of the causative disease-associated variants.

Haplotypic analysis is usually more powerful than studying individual polymorphisms in LD mapping, and haplotype-based methods have recently contributed to the identification of genes for complex diseases (Hugot et al. 2001; Ogura et al. 2001; Rioux et al. 2001). During the last decades haplotype-based analyses have successfully been used in the HLA genetics (Tuomilehto-Wolf et al. 1989; Jawaheer et al. 2002). This study used all the common variants within the *TGFB1* locus to construct haplotypes within this region. There were two haplotypes that accounted for 79% of the total number of haplotypes in this data set. This conservation of haplotypes may be attributable to the study design which only included Caucasians, or it may suggest that selective pressure acts on this functionally important locus. The occurrence of a few common haplotypes may also reflect the genome-wide pattern of haplotype blocks, where areas of conserved common haplotypes are separated by 'hot spots' of recombination and a breakdown of LD. Recently, the HapMap of chromosome 19 has been published (Phillips et al. 2003). Approximately one-third of this chromosome is contained in haplotype blocks (Phillips et al. 2003). The *TGFB1* region of the HapMap was inspected, but no haplotype block was reported in this region. However, the marker density employed has a strong impact on the ability to detect small haplotype blocks. The marker density of the HapMap in the *TGFB1* region was low and the study may have failed to detect small haplotype blocks in this region.

The diversity of this locus was assessed using the program Entropy. The promoter polymorphisms and the *TGFB1* 868 T>C variant account for most of diversity and a relatively small increase in diversity is added by the rare variants *TGFB1* 913 G>C and 1627 C>T. The promoter polymorphisms and the *TGFB1* 868 T>C variant could be used as haplotype tagging SNPs (htSNP) in the *TGFB1* region in future association studies of common complex disease, where underlying disease causing variants are likely to be of relatively high frequency. The recently identified haplotype structure in the human genome has incited a lot of effort to characterize htSNPs in candidate regions genome-wide (Daly et al. 2001; Johnson et al. 2001). These htSNPs could reduce genotyping in future LD mapping projects.

In vivo, TGF-β1 acts in combination with other peptide growth regulators, such as epidermal growth factor or its homolog, transforming growth factor alpha, platelet derived growth factor and basic fibroblast growth factor. The synthesis, secretion, activation and tissue specific expression of TGF-β1 are tightly regulated processes. Genes in the same biochemical pathways may interact and their net effect may confer the phenotype. Understanding of the role of TGF-β1 in disease pathogenesis requires dissection of these interactions and regulation of TGF-β1. TGF-β isoforms such as TGF-β2 or TGF-β3, and other related molecules such as TGF-β receptors, LTBPs or other regulatory molecules may play a role in the pathology of AS. The genes for *LTBP4* and *fibroblast growth factor 22 precursor* are located on chromosome 19 (19q13.2 and 19p13.3, respectively) and are potentially interesting positional candidate genes for AS. The *TGF-β type III receptor* gene, *betaglycan*, contains several polymorphisms (Zippert et al. 2000) which have the potential to affect TGF-β signalling. TGF-β2 mRNA has been detected in biopsy specimens from sacroiliac joints of patients with AS suggesting that TGF-β2 may play an active role in sacroiliitis (Braun et al. 1995). TGF-β2 is abundantly present in anterior chamber of the eye and could potentially be involved in the pathogenesis of AAU (Murray et al. 1999). It is also likely that as yet unknown gene-gene interactions or gene-environment interactions play a significant role in AS pathology.

Genetic studies of AS are confounded by genetic heterogeneity, uncertainty of the underlying disease model, incomplete penetrance and epistasis. When a number of genes are acting multiplicatively, the individual gene effects are likely to be modest and large sample sizes are required to capture the small effects. Animal models suggest that the potency of many susceptibility alleles is strongly dependent on the genomic context, as a result of other susceptibility alleles and suppressive modifiers (Wanstrat and Wakeland 2001). In the current study the genetic homogeneity was increased by including only the families where the proband carried the *HLA-B27* antigen in the analysis. A considerable number of families in two ethnically different populations were genotyped in order to increase the likelihood of detecting a true effect. For dominant and codominant models, the study had a relatively good power, and an association of loci with genotype relative risks of 1.4-3.1 could be detected. For recessive disease model and rare alleles, the power was poor. Because the *TGFB1* polymorphisms are located either within the coding region of the gene or in the promoter region and are likely to have functional consequences, power calculations were performed assuming that these polymorphisms would either be the actual disease-causing variants or in complete LD with them. The LD pattern around *TGFB1* gene is unknown making it difficult to predict the power of this study to detect genetic elements around the *TGFB1* locus.

Seven common SNPs and one insertion/deletion have been described within the coding region or in the promoter region of *TGFB1* gene. All the known polymorphisms in the *TGFB1* gene except the -988 C>A, +72 insertion and 713-8delC were genotyped in this study (Langdahl et al. 1997). The *TGFB1* +72 insertion was not included in the study as it exhibits almost complete LD with the *TGFB1* 913 polymorphisms (Cambien et al. 1996). The *TGFB1* -988 C>A and 713-8delC were excluded because their low frequency would make the current sample size inadequate to achieve required power (Chen and Deng 2001), and because their low frequency would make it unlikely that they contribute significantly to the observed linkage in AS. In an association study it is crucial to investigate several nearby polymorphic markers, since a disease locus may be in disequilibrium with some but not all the markers in its vicinity.

TGFB1 sequence is highly conserved between different species. There is 90% sequence homology between human, *M. musculus* and *R. norvegius Tgfb1*. There is also sequence homology between the *TGFB* isoforms, human *TGFB2* has 45% and *TGFB3* has 46% sequence homology with human *TGFB1*. Comparison of sequences from species that diverged as long as 100-300 million years ago, such as human and mouse, can be a powerful tool to identify gene-regulatory elements and protein-coding segments. Such comparisons contribute to understanding the mechanisms and history of genome evolution and determining the similarities and differences in gene organisation between the species of interest (Schwartz et al. 2000). Coding regions and noncoding regulatory regions are under selective pressure, and thus they are likely to be highly conserved between the species. Noncoding cis regulatory sequences play a central role in controlling gene expression and inter-species genomic sequence comparisons serve as a rapid and accurate means for identifying such noncoding regulatory elements. Functional characteristics of conserved noncoding sequences in the IL-4/IL-5/IL-13 cytokine cluster have demonstrated that sequence conservation is a good predictor of regulatory elements (Loots et al. 2000).

The sequence homology of the *TGFB1* genomic sequences between human and mouse was assessed using the program Pipmaker. As predicted, all the exons of the *TGFB1* gene were highly conserved. Segments of sequence homology were also noted in the promoter region and the first intron of the *TGFB1* gene, suggesting that these sequences may play a role in the regulation of the *TGFB1* expression. However, no conserved segment was noted in the region, where the promoter SNPs (*TGFB1* -800 G>A and -509 C>T) were located. Thus, this analysis does not support the role of these sequences in gene regulation.

Genes that are located within a few cM of the *TGFB1* locus and could be positional candidates in AS include *LTBP4* and several *cytochrome P450* genes. *LTBP4* lies 70.1 kb centromerically from *TGFB1* gene and immunoblot analysis of fibroblast cultures show that it binds in a covalent complex with TGF-β1-LAP and associates with extracellular matrix (Saharinen et al. 1998). Northern blot analysis detected strong expression of an approximately 5-kb LTBP4 transcript in heart, a slightly lower expression in skeletal muscle and pancreas, moderate expression in placenta and lung and weak or no expression in brain, liver and kidney (Giltay et al. 1997). *Cytochrome P450* genes in AS are discussed in chapter 6. *IL-11* is another positional candidate gene on chromosome 19, it lies 15.7 cM telomerically from *TGFB1* gene. IL-11 has pleiotropic effects on multiple tissues and its gene expression can be stimulated by IL-1α, TGF-β1 and TGF-β2 in respiratory epithelial and fibroblast cells (Du and Williams 1997).

Plasma concentrations of active and acid-activatable latent TGF-β1 are predominantly under genetic control (heritability estimate 0.54). The –509 C>T promoter polymorphism in the *TGFB1* gene has been shown to be significantly associated with higher concentrations of TGF-β1, explaining 8.2% of the additive genetic variance in TGF-β1 concentrations (Grainger et al. 1999). Furthermore, the polymorphism at positions 868 T>C and 913 G>C in the signal protein sequence of the *TGFB1* gene, which change codon 10 (Leu10Pro) and codon 25 (Arg25Pro), respectively, have been reported to be related to variations in the production of TGF-β1 (Awad et al. 1998; Yamada et al. 1998). However, it remains to be determined whether differences in the circulating concentration of TGF-β1 among individuals with different *TGFB1* genotypes affect the concentrations of the cytokine in the sacroiliac joints and entheses. The local effects of TGF-β1 in AS require further elucidation. TGF-β1 may have a promising therapeutic potential for AS, and these findings do not exclude a key role for TGF-β1 protein in pathogenesis of AS.

3.5 CONCLUSION

This study suggests that *TGFB1* polymorphisms play a minor role in AS, but do not explain the linkage of AS susceptibility to chromosome 19.

CHAPTER 4 - NOVEL HLA-DRB1 GENOTYPING METHOD BASED ON MULTIPLEX PRIMER EXTENSION REACTIONS

This chapter describes the development of a novel *HLA-DRB1* (HLA class II gene, OMIM reference number 142857) genotyping method based on multiplex primer extension reactions. The method was designed to perform high throughput *HLA-DRB1* genotyping. The advantages and limitations of the method are discussed.

4.1 INTRODUCTION

A semi-automated fluorescent method of genotyping *HLA-DRB1* alleles *HLA-DRB1*01-16* by multiplex primer extension reactions was developed and validated. This method is based on the extension of a primer that anneals immediately adjacent to the SNP with fluorescent dideoxy-NTPs (minisequencing), followed by analysis on an ABI Prism 3700 capillary electrophoresis instrument. The validity of the method was confirmed by genotyping 261 individuals using both this method and PCR-SSP or sequencing and by demonstrating Mendelian inheritance of *HLA-DRB1* alleles in families. The method provides a rapid means of performing high throughput *HLA-DRB1* genotyping using only one PCR reaction followed by three multiplex primer extension reactions.

HLA class I and class II loci located in the human MHC region on chromosome 6p21.3 encompass the most polymorphic set of genes in the human genome. With 328 alleles identified to date, *HLA-DRB1* is the most polymorphic HLA class II locus.

Extraordinary allelic diversity of these genes is believed to be maintained by selective advantage against infectious disease associated with increased heterozygosity; one allele is better able to mediate resistance to one type of pathogen, whereas the other allele is better suited to defence against a different pathogen. HLA genes play a crucial role in immune responses and are thus important for tissue transplantation, autoimmune and inflammatory diseases.

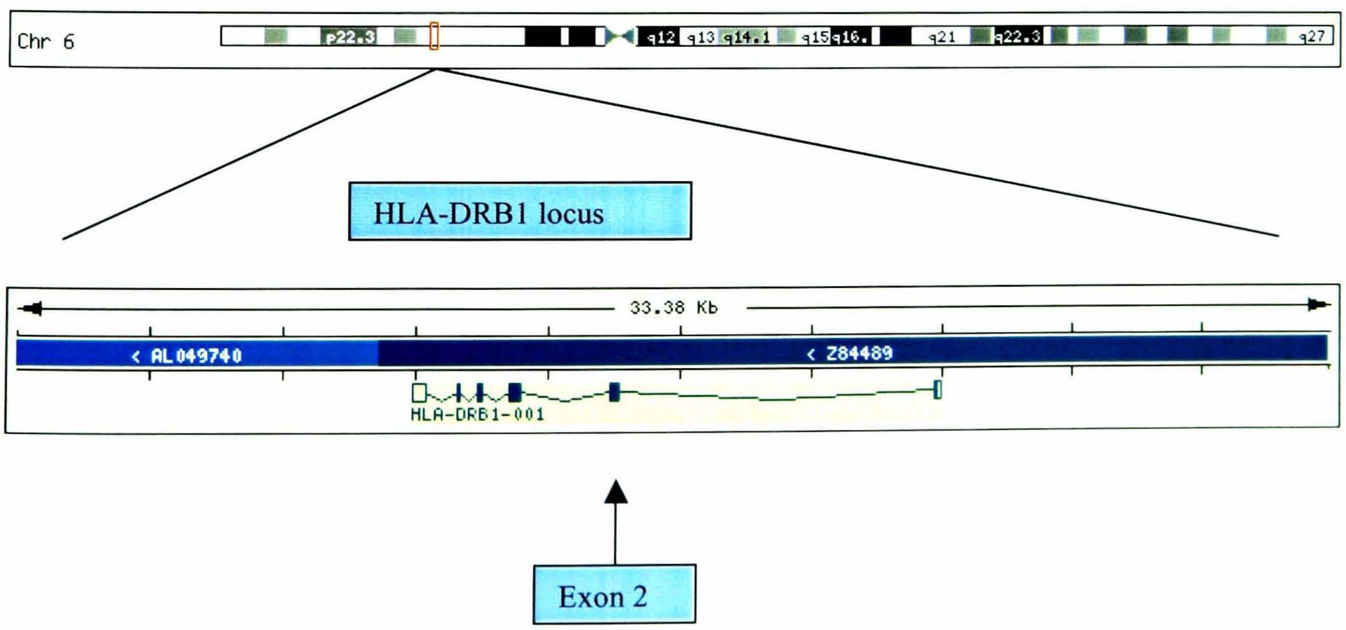
Various methods have been developed to type the polymorphisms in the HLA-class I and class II loci. These methods range from traditional serology to DNA based methods such as PCR restriction fragment length polymorphism (PCR-RFLP) (Uryu et al. 1990), PCR sequence-specific oligonucleotides (PCR-SSO) (Wordsworth et al. 1990) and PCR sequence-specific primers (PCR-SSP) (Bunce et al. 1995). Recently, various HLA-DRB1 sequencing based typing (PCR-SBT) methods have been developed to further increase the accuracy of the HLA genotyping and facilitate the identification of novel *HLA-DRB1* polymorphisms (Kotsch et al. 1999; Sayer et al. 2001). *HLA-DRB1* genotyping methods based on fluorescence-marked sequence-specific priming (Taqman assay) have been developed to increase automation and throughput (Tremmel et al. 1999; Slateva et al. 2001).

Rapid and accurate genotyping of polymorphisms is of significant scientific importance in studies of linkage and association with disease. Unravelling the genetic basis of complex diseases requires the genotyping of a large number of samples, which is facilitated by high throughput genotyping methods. The vast majority of polymorphisms in the *HLA-DRB1* locus are located within the second exon of the gene, which makes it amenable to the multiplex primer extension approach. Figure 4.1

represent the structure of the *HLA-DRB1* gene. The feasibility of the primer extension technique for HLA typing was demonstrated by Pastinen et al. for *HLA-DQA1* typing and *HLA-DRB1*02* subtyping (Pastinen et al. 1996). Here a high throughput genotyping method for the classic *HLA-DRB1* alleles **01-16* based on multiplex primer extension reactions is described. This is achieved by the use of the commercially available ABI Prism® SNaPshot™ technology (Applied Biosystems, Foster City, CA, USA). SNaPshot™ is a minisequencing method that is based on the single nucleotide extension of a primer that terminates immediately adjacent to the SNP using fluorescently labelled ddNTPs.

Figure 4.1 The location and structure of the *HLA-DRB1* gene.

HLA-DRB1 gene is located on the short arm of chromosome 6 within the 6p21.3 region. It is part of the human MHC, or HLA region, and it is one of the most variable loci known in the human genome. *HLA-DRB1* consists of six exons. The polymorphisms that characterise the vast majority of *HLA-DRB1* alleles are located within the second exon of the gene. The second exon is extremely polymorphic and encodes part of the antigen-binding groove. The polymorphisms in the second exon tend to be restricted to specific regions of variability. These variable regions are interspersed between regions that are relatively conserved between all HLA class II genes. The PCR amplification primers used in this method were designed in the conserved region of the gene flanking the hypervariable regions. The figures depicting the chromosome 6 and the structure of the *HLA-DRB1* gene were obtained from the Ensembl webserver at www.ensembl.org.



4.2 SUBJECTS AND METHODS

4.2.1 Subjects and families

Two hundred and sixty-one probands of Finnish families affected with AS were initially genotyped using direct sequencing and PCR-SSP (Bunce et al. 1995). The accuracy of the SNaPshot^{TM} genotypes was assessed by regenotyping these probands by the SNaPshot^{TM} protocol. The remaining family members [n=412] were genotyped using the SNaPshot^{TM} protocol, and Mendelian inheritance of *HLA-DRB1* alleles checked using the GAS version 2.0 software [A. Young, unpublished].

4.2.2 DNA amplification

A 275 base pair (bp) fragment from exon 2 of all the *DRB* genes was amplified using the primers DRBforward 5' CCCCACAGCACGTTTC 3' and DRBreverse 5' GCCGATGCACTGTGAAGCTCTC 3'. Figure 4.2 represents positive reactions for the PCR (PCR product one). This PCR product was used to perform high-throughput genotyping of *HLA-DRB1* alleles **01-14*.

The multiplex primer extension reactions performed using the PCR product one do not resolve the *HLA-DRB1*02* subtypes, *HLA-DRB1*15* and **16*. A separate PCR and primer extension reaction was designed and applied to some of the samples to perform *HLA-DRB1*15/16* subtyping. This PCR reaction specifically amplified all *HLA-DRB1*02* subtypes using a forward primer 5' CACGTTTCCTGTGGCAGCCTAAGA 3' and the same generic DRBreverse primer (PCR product two) (Figure 4.3).

The amplification was performed in a 20 μl reaction mixture containing 40-100 ng DNA, 10 mM Tris-HCl (pH 8.3 at 25°C), 50 mM KCl, 1.5 mM $MgCl_2$, 50 μM of each dNTP, 0.25 μM of each primer and 1 unit of Taq Gold DNA polymerase. The cycling conditions were as follows: 95°C for 10 min, 10 cycles of 94°C for 30 s, 60°C (-0.5°C per cycle) for 40 s, 72°C for 20 s, 30 cycles of 94°C for 30 s, 53°C for 40 s, 72°C for 20 s and 72°C for 1 min.

PCR products (5 μl) were separated on a 2% agarose gel stained with ethidium bromide, and visualised under ultraviolet light. The remaining PCR products were incubated with 1.5 units of SAP and 1 unit of Exo I at 37°C for at least one hour to degrade unincorporated primers and dNTPs, 75°C for 15 min to denature the enzyme and 4°C for 15 min. Reduced amount of dNTPs and primers were used in the PCR reaction in order to minimise the use of SAP and Exo I.

Figure 4.2 2% agarose gel image of PCR amplification of pan *HLA-DRB* genes (*HLA-DRB1* and *DRB2-9*).

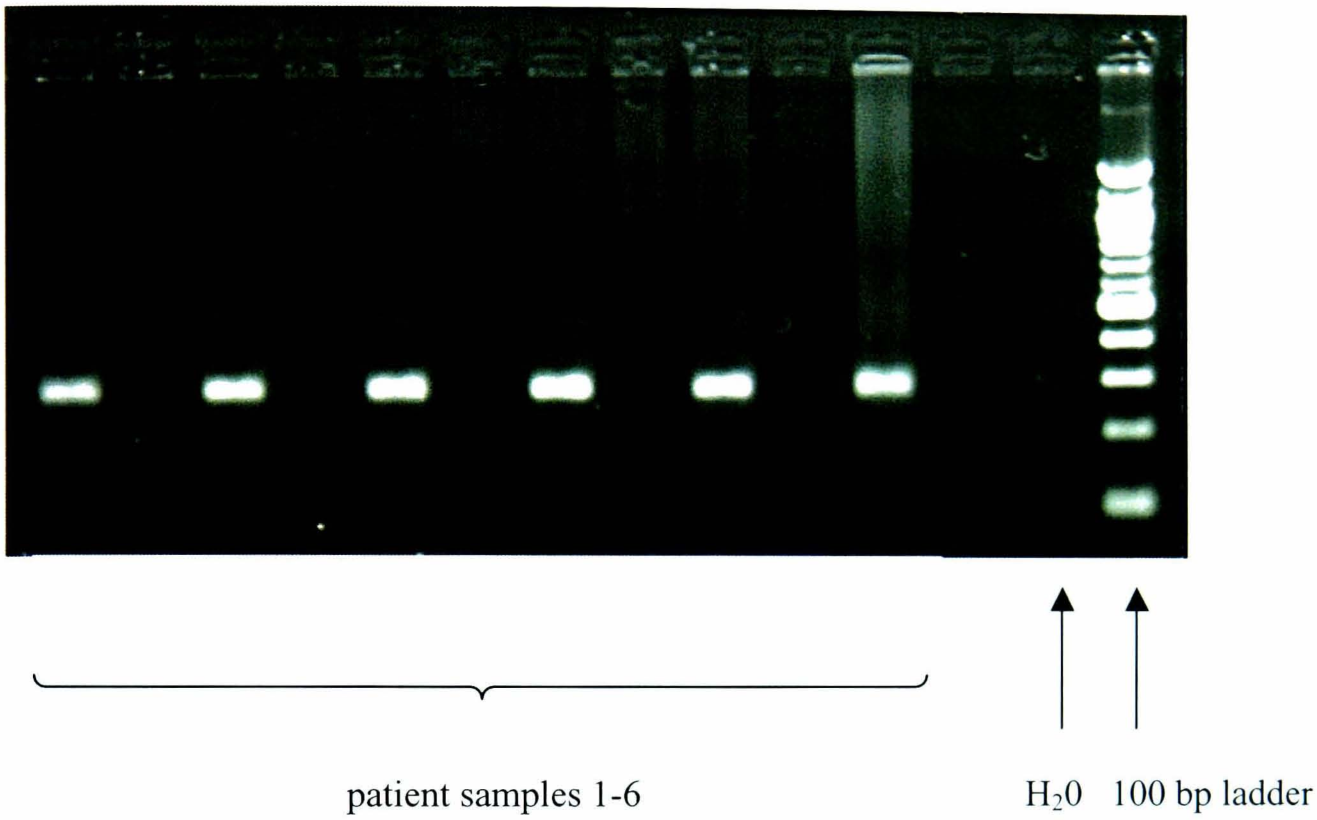
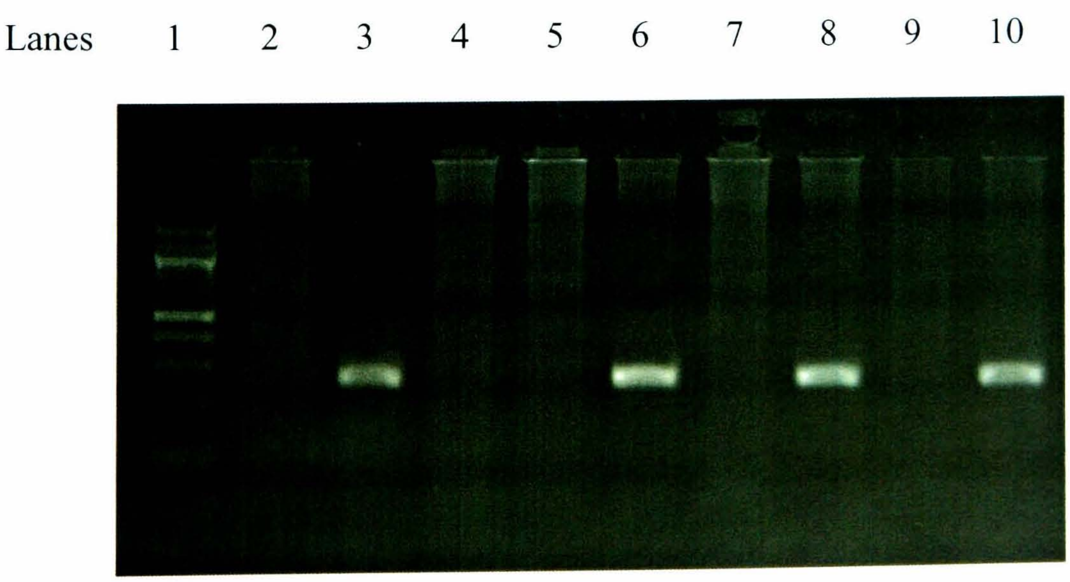


Figure 4.3 2% agarose gel image of PCR amplification of *HLA-DRB1*15/16* alleles.

Lane 1 – 100 bp ladder; Lanes 3, 6, 8 and 10 – *HLA-DRB1*15/16* positive samples; Lanes 2, 4, 5, 7 and 9 – *HLA-DRB1*15/16* negative samples



4.3 RESULTS

4.3.1 Primer optimisation

The primers were designed using the IMGT/HLA Database as a reference sequence (Robinson et al. 2003). This is a specialist database for sequences of the genes within the MHC region and includes all the HLA sequences officially recognised and named by the WHO Nomenclature Committee for Factors of the HLA System. Oligo software version 4.0 was used to analyse the PCR and extension primers [www.oligo.net]. The melting temperatures of the primers varied between 49-69 °C and the primers did not contain stable secondary loops or 3' primer complementarity. SNP variants specified by the extension primers were carefully selected in order to avoid false positive extension from other *HLA-DRB1* alleles, other *HLA-DRB* genes (*HLA-DRB3, B4* and *B5*) or sequence-related pseudogenes (*HLA-DRB2, B6-B9*).

The extension primers were designed either on the sense or antisense DNA strand adjacent to the SNPs of interest. The primer extension reactions were performed in three multiplexes to avoid overlapping extension primers where SNPs are very close together. The length of the extension primers varied between 16 and 42 nucleotides so that the products could be separated by capillary electrophoresis. Between 16 and 27 nucleotides of each primer was sequence specific, with poly A added to increase the primer length where appropriate. The extension primers in the SNaPshot^{TM} multiplexes were determined in a way that multiple peaks of the same nucleotide of fluorescent dye did not appear next to each other when electrophoresed. The final development of primers included thirteen different extension primers that were optimised and validated for the full *HLA-DRB1* typing (*HLA-DRB1*01-16*). Table 4.1 describes the single nucleotide polymorphisms specified by the optimised set of extension primers. The sequence, length, melting temperature (T_m) and localisation of these extension primers are listed in table 4.2.

4.3.2 Primer extension reactions and demonstration of the method

PCR product one was used for three subsequent multiplex SNaPshot™ reactions, PCR product two was used for a uniplex SNaPshot™ reaction. *HLA-DRB1* alleles **01-14* were genotyped from PCR product one. Three multiplex primer extension reactions were designed for these alleles in order to avoid overlapping extension primers where SNPs are very close together. The PCR product two was used to perform *HLA-DRB1*15* and *HLA-DRB1*16* subtyping in some of the samples.

Figure 4.4 illustrates the positions of the extension primers and the polymorphisms specified by them. Figure 4.5 illustrates examples of positive primer extension reactions for the different extension primers.

The primer extension reactions were performed in a 10 µl reaction mixture containing 3 µl of the purified PCR product, 0.5 µl of SNaPshot™ reaction mixture per SNP and 0.18 µM (for *HLA-DRB1*04, *07, *09*) or 0.36 µM of each extension primer. The primer extension cycling conditions were as follows: 25 cycles of 96 °C for 10 s, 50 °C 5 s, 60 °C 30 s and 4 °C for 15 min. To prevent unincorporated fluorescent ddNTPs to co-migrate with the primer extension products during electrophoresis, the reactions were treated with 1 U SAP at 37 °C for 1 hour followed by 15 min at 75°C and 15 min at 4°C. Aliquots of 3 µl of SNaPshot™ product, 0.5 µl of GeneScan -120 LIZ Size Standard and 6.5 µl of Hi-Di formamide were combined in a 96-well Thermo-Fast®, Non-skirted plate (ABgene, Surrey, UK). The products were separated using POP5 polymer and detected by capillary electrophoresis on an ABI Prism 3700 DNA Analyzer (Applied Biosystems, Warrington, UK). Products were then sized and analysed using Genotyper® version 3.5 (PE Biosystems). Genotypes were semi-automatically assigned and checked manually by comparing with the positive control samples included in each plate.

Whilst this method is simple and involves few reactions, there are some *HLA-DRB1* combinations which it does not fully resolve. Because the extension primers for *HLA-DRB1*03* or *HLA-DRB1*08* alleles are not specific but also detect other alleles (see table 4.1), the method does not resolve the following heterozygote combinations of *HLA-DRB1* alleles: *HLA-DRB1*03/11* or *HLA-DRB1*11/11*; *HLA-DRB1*03/13* or *HLA-DRB1*13/13*; *HLA-DRB1*03/14* or *HLA-DRB1*14/14*, *HLA-DRB1*08/12* or *HLA-DRB1*12/12*. Similarly, because *HLA-DRB1*13* alleles and *HLA-DRB1*01* subtype *HLA-DRB1*0103* are specified by the same extension primer, number 5, the method does not distinguish the *HLA-DRB1*0103/13* genotype from those who are *HLA-DRB1*0103* homozygotes. These ambiguous combinations occurred with 6.9% frequency in this data set. They are easily resolved by a single PCR reaction using PCR-SSP, specific either for *HLA-DRB1*03, HLA-DRB1*08* or *HLA-DRB1*0103* depending on the ambiguity concerned (see (Bunce et al. 1995) and chapter 5).

Table 4.1 Description of the SNPs specified by the extension primers.

The determination of the allele can be deducted according to the specific nucleotide, which is extended by the extension primers.

PRIMERS				
MULTIPLEX I	**CODON***	**HLA-DRB1 SUBTYPES SPECIFIED BY THE EXTENSION PRIMERS**		
1	43/3	**A**: DRB1*010101-0109	No extension or **G**: all the other	
2	55/3	T: DRB1*09012-0902	No extension or **C**: all the other	
3	59/1	**G**: DRB1*100101-100102	No extension or **C**: all the other	
4	82/3	**A**:DRB1*070101-0707	**G**: all the other	
MULTIPLEX II				
5	100/2	**A**:DRB1*0103,0402,0414,0437,1102-3,1111,1114,1116,1120-1,1136,1140-1,130101-202,1304,1308,1315-7,1319-20,1322-4,1327-9,1331-2,1334-6,1338-41,1343,1345,1348,1351-3,1416,1510	**C**:DRB1*0106,0109,150101-9,1511-3,1309,1424,1437,DRB5*0106,0202-4	No extension: all the other
6	41/2	**G**:DRB1*030101-23,0820,110101-110402,1106-7,111202-1121,1123-1129,1131-43,130101-1305,130701-1316,1318-1353,140101-1403,140501-1409,1413-1414,1416-1427,1429-1430,1432-1445	**C**:DRB1*0814, A:DRB1*0821	No extension or **T**:all the other
7	87/2	**A**:DRB1*0308,0415,110101-1143,1204,1411	**C**: all the other	
8	42/3	**G**:DRB1*150101-201, 150203, 1503-13, 160101-1608	No extension or **T**:all the other	
9	62/1	**C**:DRB1*040101-0444	**A**: all the other	
10	55/3	**A**:DRB1*120101-1207	**G**:DRB1*010101-0109,100101-100102	No extension or **C**:all the other
MULTIPLEX III				
11	45/1	**T**:DRB1*080101-080401,080403-0819,0821-24,120101-1208,1317,1404,1411,1415,1428,1431	**C**: all the other	
12	86/2	**G**:DRB1*0310,0706,0808,0811,1343,1345,140101-2,1404,140701-2,1410,1416,1422,1425-6,1428,1431-2,1435,1438-9,DRB3*0216	No extension or **T** or **A**:all the other	
DRB1*15/16				
13	76/2	**T**:DRB1*150101-6,1508-10,1512-13	**A**:DRB1*160101-08	

* The reference sequence used to determine the codons was NM_002124.

Table 4.2 Description of the extension primers.

PRIMER	SEQUENCE 5'-3' *	PRIMER	ORIENTATION	Tm** (°C)	LOCATION
NUMBER		LENGTH			EXON2***
MULTIPLEX I		base			
1	GGCAGC**C**TAAGTTTGA	16	Sense	49.2	13-28
2	TCAACGG**T**ACGGAGCGGGTGCGGTA	25	Sense	69.5	40-64
3	*AAAAAAAAAAAA*CCTCTTGGTTATGGACGC	30	Antisense	61.3	76-93
4	*AAAAAAAAAAAAAAAAAAAAAAA*CGGGCGGTGACGGAGCT	40	Sense	66.5	129-145
MULTIPLEX II					
5	GGACATCCTGGAAGACG	17	Sense	55.2	182-198
6	*AAAA*AAGAAATGACACTCAGAC	22	Antisense	52.8	23-40
7	*AAA*TGACGGAGCTGGGGCG**A**CCTGATG	27	Sense	69.5	136-159
8	*AAAAAAAAAAAAAAAA*TGTGGCAGCCTAAGAG	32	Sense	60.5	10-25
9	*AAAAAAAAAAAAAAAAAA*CCTGGACAGATACTTCTAT	37	Sense	60.6	65-83
10	*AAAAAAAAAAAAAAAA*TTCAATGGGACGGAGCGGGTGCGGTT	42	Sense	69.4	39-64
MULTIPLEX III					
11	AGTACTCTACGGGTGAGTGT	20	Sense	57.3	13-32
12	*AAAAAAA*GCTGTTCCAGTG**T**TCCGCA	26	Antisense	61.6	158-176
DRB1*15/16					
13	ACAGCGA**A**GTGGGGGAGT	18	Sense	58.2	109-126

* Mismatches introduced to destabilize stable secondary hairpin formations are in bold and underlined. Poly A tails are in italics.

** T_m of the primers was calculated without the poly A tail.

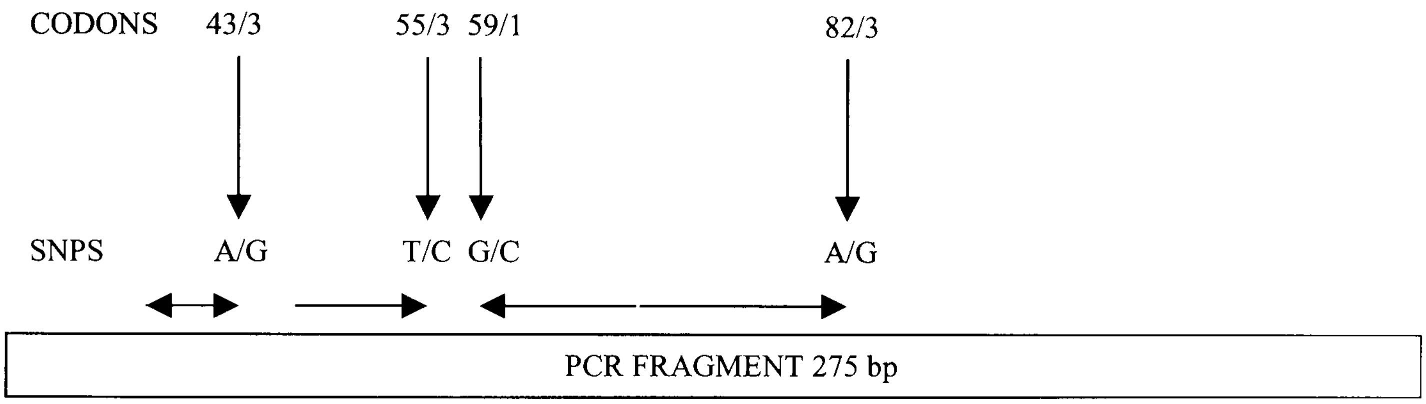
*** Location of the primers is given in nucleotides. The first nucleotide denotes the first nucleotide of the second exon of the *HLA-DRB1* gene. (The ID of the Ensembl reference sequence used is OTTHUMT00006012178.)

Figure 4.4 Illustration of the positions of the extension primers and the polymorphisms specified by them.

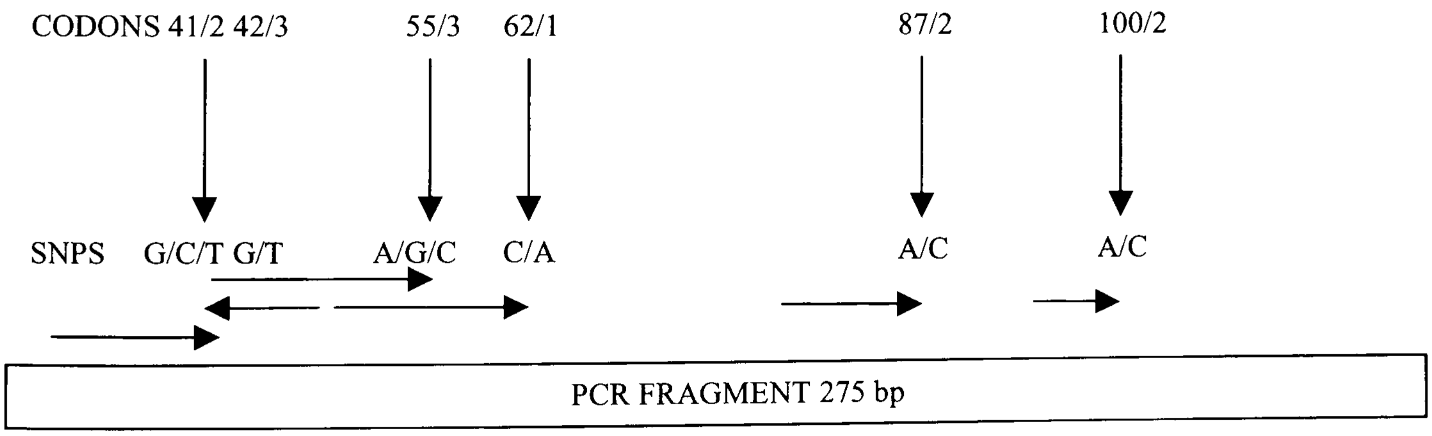
The figure demonstrates the multiplex reactions I, II and III performed using the PCR product one and uniplex reaction of *HLA-DRB1*15/16* using the PCR product two. In multiplexes I and II some poly A tails overlap, but the remaining allele specific part of the primer sequences do not overlap by more than 3 bases.

PCR PRODUCT ONE:

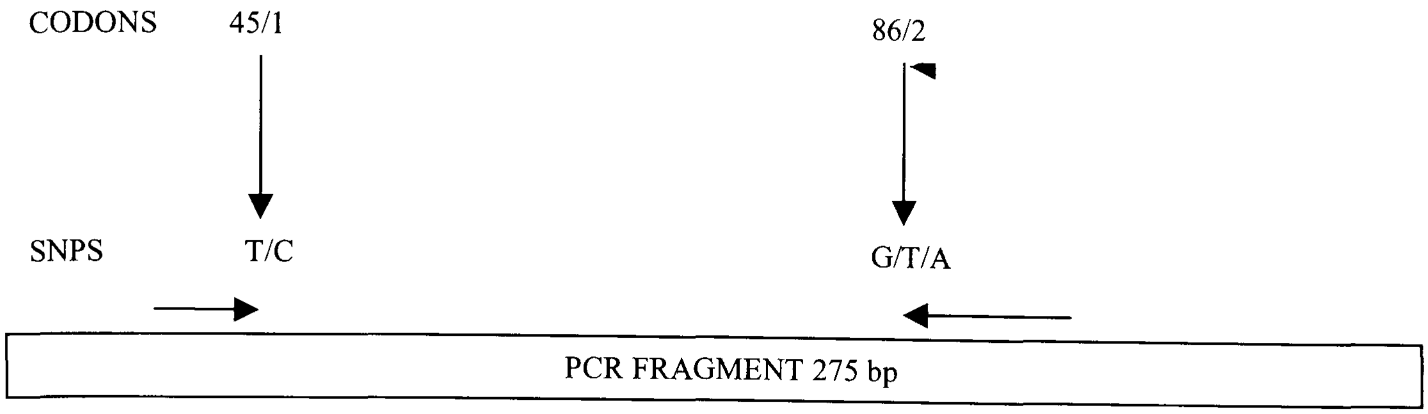
MULTIPLEX I



MULTIPLEX II



MULTIPLEX III



PCR PRODUCT TWO:

UNIPLEX REACTION

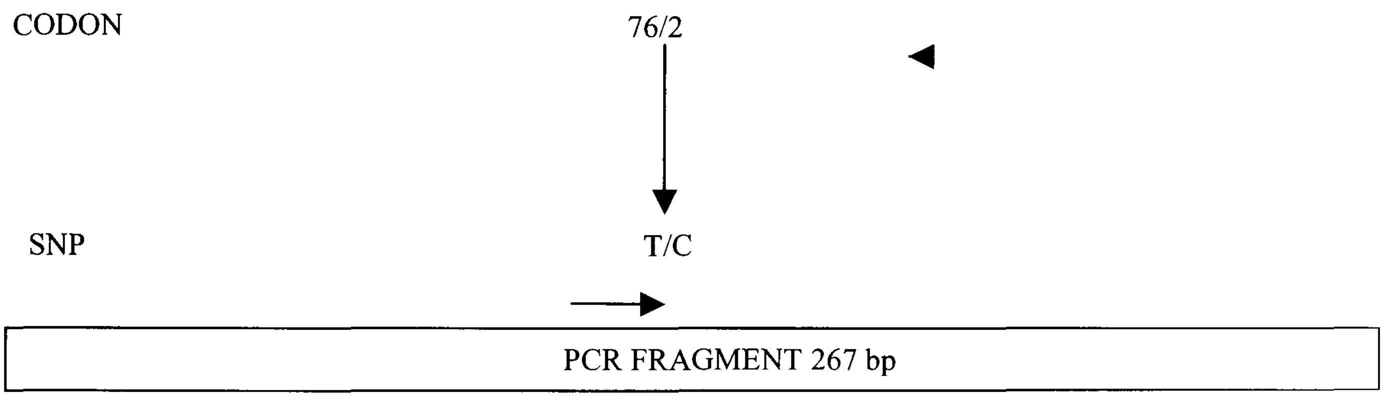
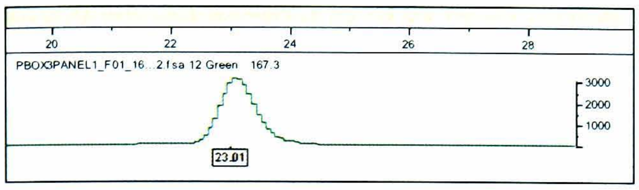


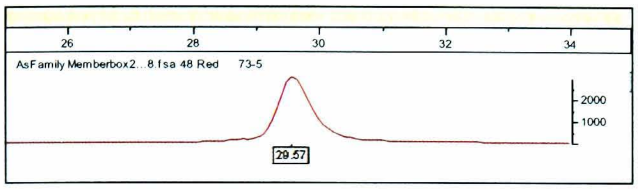
Figure 4.5 The figure illustrates positive primer extension reactions for the different extension primers in multiplexes 1-3 and in the uniplex reaction of *HLA-DRB1*15/16*.

MULTIPLEX I

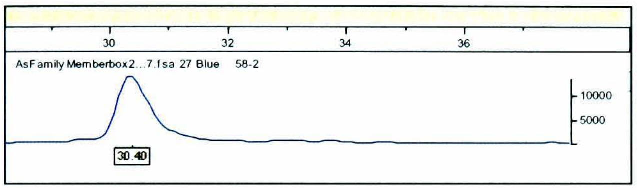
Primer no 1 - DRB1*01 positive sample



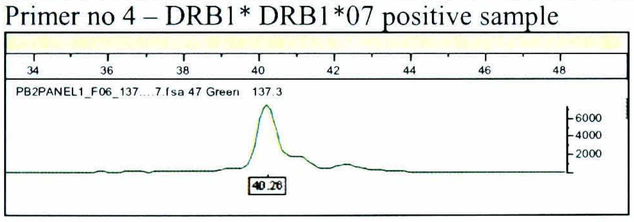
Primer no 2 – DRB1*09 positive sample



Primer no 3 – DRB1*10 positive sample

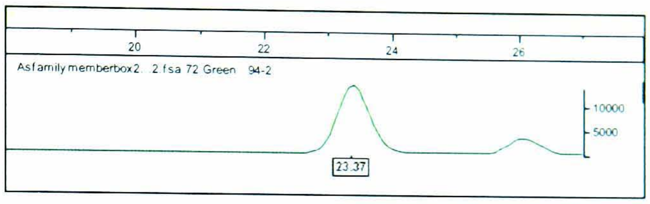


Primer no 4 – DRB1* DRB1*07 positive sample

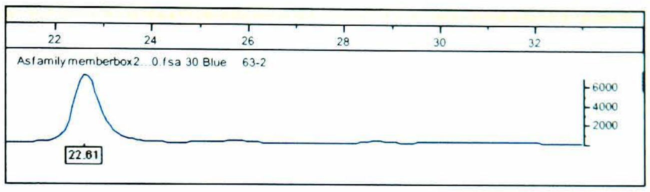


MULTIPLEX II

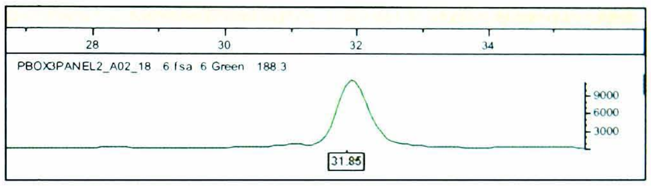
Primer no 5 - DRB1*13 positive sample



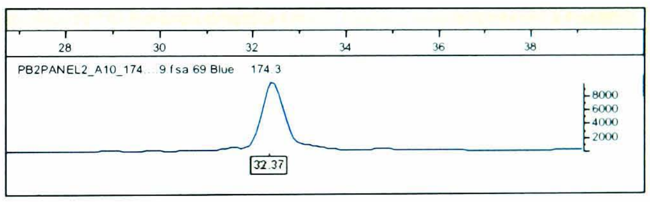
Primer no 6 – DRB1*03/08/11/12/13/14 positive sample



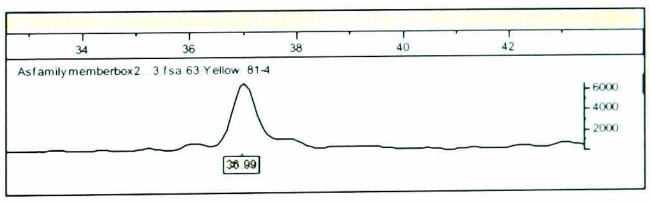
Primer no 7 - DRB1*11 positive sample



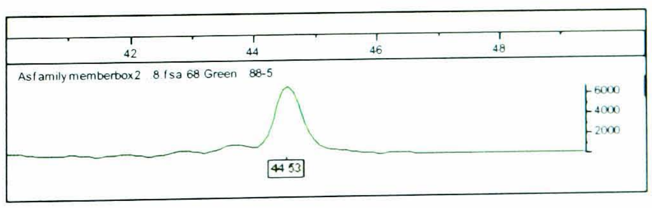
Primer no 8 – DRB1*02 positive sample



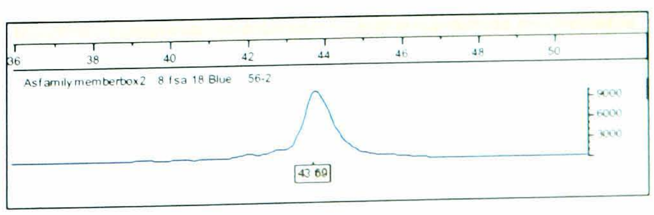
Primer no 9 – DRB1*04 positive sample

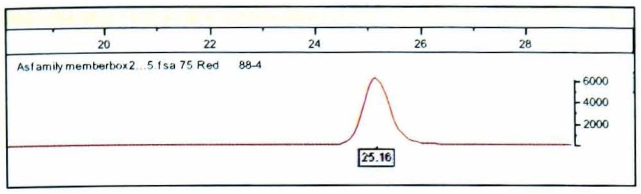
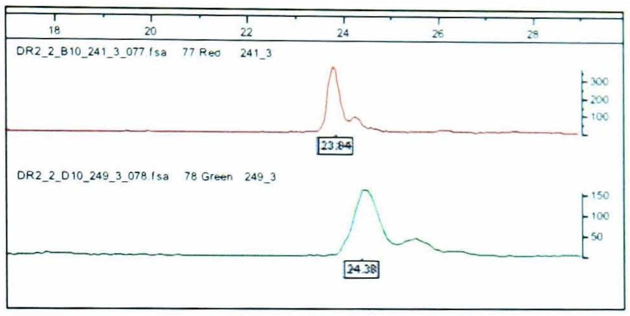
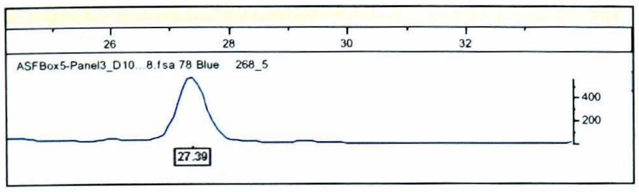


Primer no 10 – DRB1*12 positive sample



Primer no 10 – DRB1*01/10 positive sample



<table>
<tr><th>MULTIPLEX III</th><th>DRB1*15/16</th></tr>
<tr><td>Primer no 11 – DRB1*08/12 positive sample

</td><td rowspan="2">Primer no 13 - DRB1*15 (red colour) and DRB1*16 (green colour) positive samples.

</td></tr>
<tr><td>Primer no 12 – DRB1*14 positive sample

</td></tr>
</table>

4.3.3 Validation

In addition to confirming Mendelian inheritance of genotypes, the accuracy of the method was validated by genotyping 261 probands using both this method and sequencing or PCR-SSP (Bunce et al. 1995), a method that is widely used for histocompatibility testing. Without any repetition of genotyping of individual samples, there were 17 allelic errors out of 522 alleles genotyped with SNaPshot™ (3.3%). All the errors were due to allele drop out resulting in erroneous scoring of a heterozygote person as a homozygote. Ten individuals resulted in a failed reaction after the first genotyping (success rate 97%). All the samples that failed and all the apparent homozygote samples were regenotyped with the SNaPshot™. After regenotyping, all the individuals were consistently genotyped both with the SNaPshot™ approach and the PCR-SSP (success rate 100%). All the genotypes were consistent with the genotypes obtained by the previous PCR-SSP and direct sequencing giving a zero overall error rate.

Some of the consistently genotyped individuals were further used as positive controls in subsequent genotyping plates. A positive control of each of the classic *HLA-DRB1* allele **01-16* were included in each plate and the positive controls were used in manual checking of the semi-automatically assigned genotypes. A negative H_2O control was included in each plate in order to check that the PCR was not contaminated.

4.4 DISCUSSION

The MHC region is of primary scientific importance in relation to autoimmunity and infection. The recently completed sequence map of the MHC and the identification of numerous novel genes within this region contributed to increased interest in high resolution MHC mapping studies (1999). High levels of polymorphism, linkage disequilibrium and gene clustering confound the precise mapping of contributory genes within the MHC (Allcock et al. 2002). Several complex diseases, such as rheumatoid arthritis and type 1 diabetes, exhibit allelic and locus heterogeneity within this region [Okamoto, 2002 #243; Lie, 1999 #242], which adds an additional layer of complexity to fine-mapping studies of the MHC region. Rapid high-throughput genotyping methods are required to facilitate the dissection of complex genetic associations within the MHC region. The method described here is cost-effective, rapid, accurate, sensitive, robust and amenable to high-throughput genotyping. It is based on a widely accessible technology. It is technically relatively easy to carry out and could be performed in most modern genetics laboratories. This study describes a novel approach for genotyping of *HLA-DRB1* alleles *HLA-DRB1*01-14* from only one PCR product by multiplex primer extension reactions. It is especially suitable for disease association studies, where the amount of DNA is not sufficient for numerous PCR amplifications particularly if high resolution genotyping is not required. The method involves considerably fewer PCR reactions than PCR-SSP in particular, reducing the time required, DNA usage and expense.

This method as described above does not perform *HLA-DRB1* high resolution subtyping but there is no theoretical reason why primer extension methods could not be developed for that application. Indeed a method for *HLA-DRB1*02* subtyping using primer extension has been reported (Pastinen et al. 1996). High resolution HLA typing is of more importance for tissue typing prior to organ transplantation (Muro et al. 2002) and this method would require further development for these purposes. However, it is useful for screening of subjects for whom expensive and time consuming high resolution genotyping is not necessary.

Oligonucleotides with a mismatched 3' –residue will not function as primers in the PCR reaction under appropriate conditions (Newton et al. 1989). This principle can be extended to the extension primers in the primer extension reactions. This fact was exploited in the oligonucleotide design to ensure that some of the extension primers have a mismatched 3' –residue in the other *HLA-DRB* genes or the other *HLA-DRB1* alleles, so that they are not specified by the extension primer. This increases the specificity of the primer extension reaction but has the disadvantage is that there is no negative control peak and thus false negative allele scoring is possible. The primer extension reactions for all the homozygous samples were repeated in order to detect possible false negative scoring.

Initial amplification of all the *HLA-DRB* genes can cause erroneous products on the primer extension reactions, interfere with the data analysis and cause background noise. Additionally, it greatly limits the possibilities of the extension primer design. Additional peaks were observed appearing in the data set. These peaks were generally shorter than 20 nucleotides and could be distinguished from the true alleles by comparing the size and height of the products with positive controls. The intensity of the fluorescence of the spurious peaks was generally considerably lower than that of the true peaks. Positive controls for all the alleles were used in each plate. Scoring of the *HLA-DRB1* genotypes was performed semi-automatically by the Genotyper® software and checked manually by comparing the location and the size of the peaks with the positive control samples with known *HLA-DRB1* types.

Different peak heights were observed in the multiplex SNaPshot™ reactions of primer extension signals for different polymorphisms and samples. Possible explanations include differential PCR amplification of alleles, differential annealing of extension primers, differential efficiencies of the incorporation of the ddNTPs for each allele-specific primer extension reaction and unequal emission energies of different fluorescent dyes (Norton et al. 2002). Linear relationship between primer concentration and primer extension signal strength has been reported (Norton et al. 2002). Manual comparison of peak heights between samples and positive controls were performed and peaks of at least 50% of the height of the positive control were accepted. All the positive control samples used were heterozygotes for *HLA-DRB1*.

The existence of other *HLA-DRB* genes and sequence-related pseudogenes has made it a challenging task to develop accurate and sensitive *HLA-DRB1* genotyping methods suitable for high throughput genotyping. The advantage of this approach is that it allows the use of only one PCR reaction to amplify the classic *HLA-DRB1*01-14* alleles. The disadvantage is that the PCR primers used also amplify the other *HLA-DRB* genes (*HLA-DRB3, B4* and *B5*) and sequence-related pseudogenes (*HLA-DRB2* and *B6-B9*). No polymorphisms specific for all the *HLA-DRB1* alleles but not for any of the related *HLA-DRB* genes have been recognised, despite extensive sequencing of introns flanking exon 2 (Kotsch et al. 1999). Therefore, *HLA-DRB1* PCR-SBT methods typically employ several group-specific PCR primers to obtain a product specific for only *HLA-DRB1* alleles. This product can be subsequently sequenced to determine the exact subtype. A similar approach using group-specific *HLA-DRB1* PCR reactions prior to primer extension reactions may yield more accurate and specific *HLA-DRB1* genotyping results in the future.

4.5 CONCLUSIONS

In conclusion, this method has advantages over other *HLA-DRB1* genotyping methods in sample throughput and speed, but is not designed to replace traditional high resolution typing methods used in clinical settings. The alleles can be semi-automatically assigned, which increases the sample throughput in comparison with PCR-SBT and PCR-SSO. The method is less time consuming than PCR-SBT. The main limitation of this method is its low resolution, which could be overcome by employing a group of *HLA-DRB1* specific PCR primers. Genotyping highly polymorphic HLA genes is of primary interest in MHC studies aiming to refine the genetic associations with the disease. Deciphering the genetic associations within the MHC in common complex genetic diseases requires large samples sizes to be genotyped in order to achieve adequate power. A novel way of performing high throughput *HLA-DRB1* typing was demonstrated, which will facilitate future genetic studies within the MHC region.

CHAPTER 5 - THE EFFECT OF HLA GENES IN SUSCEPTIBILITY TO AND CLINICAL CHARACTERISTICS OF AS IN A FINNISH POPULATION

This chapter presents the results of a study designed to investigate the influence of *HLA-DRB1* and *HLA-B27* alleles and *HLA-DRB1-B27* haplotypes in susceptibility to, and severity of, AS in a Finnish population.

5.1 INTRODUCTION

Previous studies have implicated several distinct *HLA-DRB1* associations in AS, but only the association between *HLA-DRB1*01* and disease susceptibility has been reported by several groups. In a large British AS cohort, an association with *HLA-DRB1*01* was noted with susceptibility to the disease (Brown et al. 1998a). A Mexican study also reported a significant increase of *HLA-DRB1*01* in cases with spondyloarthropathies (Vargas-Alarcon et al. 2002b). A small Azorean study reported an association between *HLA-DRB1*01* and *HLA-B27* negative spondyloarthritis (p=0.0009) (Bruges Armas et al. 2000b), but not with *HLA-B27* positive disease (Bruges Armas et al. 2000a). In contrast, no significant class II associations were found in a previous Finnish study (Westman et al. 1996). A recent Sardinian study reported an association between an *HLA-DRB1*15-B27* haplotype and AS (La Nasa et al. 1993; Fiorillo et al. 2003). A French study noted that *HLA-DRB1*04* allele was transmitted in excess to cases with spondyloarthropathy, but this was not independent of linkage to *HLA-B27* (Said-Nahal et al. 2002). *HLA-DRB1*04* is also associated with the development of radiographically evident sacroiliitis in patients with juvenile idiopathic arthritis (Flato et al. 2002).

Evidence that *HLA-DRB1* alleles might modulate the clinical expression of AS comes from studies of AS complicated by AAU, peripheral arthritis or IBD, and from juvenile onset AS. An association between the *HLA-DRB1*08-DPB1*0301* haplotype and juvenile AS was noted in a Norwegian population (Ploski et al. 1995). In the Japanese, an association between *HLA-DRB1*08* and AS complicated by AAU has been reported (Monowarul Islam et al. 1995). However, in this study the patients with history of AAU had a considerably lower mean age of onset than the remaining patients, suggesting that AAU is associated with an early onset disease (Ploski et al. 1996). *HLA-DRB1*04* and **07* are reported to be associated with AS complicated by peripheral arthritis (Armstrong et al. 1983; Aaron et al. 1985; Miehle et al. 1985; Sanmarti et al. 1987). The *HLA-DRB1*01* subtype, *HLA-DRB1*0103*, is associated with both isolated IBD and AS complicated by IBD (Stokkers et al. 1999; Laval et al. 2000).

Taken together these data suggest that either HLA class II genes themselves or genes on an *HLA-B27-DRB1* haplotype act as susceptibility factors to AS. Alternatively, they may modify the clinical manifestations of the disease in the context of the genetic background of the population. The Finnish population has a rather restricted gene pool (Siren et al. 1996), a property which may be useful in dissecting the relevant disease causing polymorphisms (Varilo et al. 1996). The results of a study of the *HLA-DRB1* and *HLA-B27* alleles and haplotypes in susceptibility to, and severity of, AS in a Finnish population are presented in this chapter.

5.2 SUBJECTS AND METHODS

5.2.1 AS families and controls

AS families. Fifty-three sporadic AS cases and 620 individuals from 208 AS families were recruited to the study from the Rheumatism Foundation Hospital in Heinola, Finland.

Control haplotypes. One thousand two hundred and fifty-four (105 HLA-B27-positive) haplotypes of Finnish parents of children with insulin-dependent diabetes mellitus (IDDM) not transmitted to the diabetic offspring, nor found in siblings or parents with IDDM, were used as a control sample set (Tuomilehto-Wolf et al. 1989).

5.2.2 *HLA-DRB1* genotyping

HLA-DRB1 genotyping was performed using either direct sequencing, the primer extension method developed in this laboratory or the PCR-SSP approach (Bunce et al. 1995). Positive and negative controls were used in all the reactions. One hundred and twenty-eight probands were genotyped using the direct sequencing method. This method was validated by regenotyping 44 samples previously genotyped by sequencing using the PCR-SSP approach (Bunce et al. 1995). The rest of the probands [n=133] were genotyped using the PCR-SSP approach. A novel *HLA-DRB1* genotyping method based on primer extension reactions was developed and is described in detail in chapter 4. This method was validated by regenotyping all the probands previously genotyped by sequencing and PCR-SSP using this method [n=261]. The primer extension method was further used to genotype all the family members [n=412], Mendelian segregation of *HLA-DRB1* alleles was confirmed. The control samples were typed for the *HLA-DRB1* alleles **01* to **10* using serological methods by Eva Tuomilehto-Wolf and colleagues (Tuomilehto-Wolf et al. 1989).

5.2.2.1 Sequence based method

The sequence based method involved 8 group-specific PCR amplifications with the primers HLA-DRB1*01 (5'CAGTGTCTTCTCAGGTGGCT), HLA-DRB1*15/16 (5'GGCCGCCTTGTGACCGGATG), HLA-DRB1*03/08/11/13/14 (5'GCCTCAGGAAGACAGAGGAG), HLA-DRB1*04 (5'CTTGGGATCAGAGGTAGATTTT), HLA-DRB1*07 (5'CGGCGTCGCTGTCAGTGTT), HLA-DRB1*09 (5'CAGTTAAGGTTCCAGTGCCA), HLA-DRB1*10 (5'CCCACAGCGTTCTTGGAGG) and HLA-DRB1*12 (5'AGTGTCTTCTCAGGACGCCA) prepared in a 50:50 mix with a generic reverse primer with an M13-21 sequencing tag (5'TGTAAAACGACGGCCAGTGCCGCTGCACTGTGAAGCTCTC). Amplification was performed in a 10-μl reaction mixture containing 50 ng DNA, 10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$ (for HLA-DRB1*01 and HLA-DRB1*03/08/11/13/14) or 2.5 mM $MgCl_2$, 25 μM of each dNTP, 0.25 U AmpliTaq Gold polymerase (Applied Biosystems, Warrington, UK) and 0.4 μM primers. The cycling conditions were as follows: 94°C for 14 min; 35 cycles of 94°C for 30 s, 55°C (for HLA-DRB1*15/16, HLA-DRB1*04 and HLA-DRB1*10) or 60°C for 30 s and 72°C for 30s. PCR products were separated on a 3% agarose gel stained with ethidium bromide, and visualised under ultraviolet light. PCR products were subsequently sequenced using M13-21 Big Dye dye-primer sequencing kit (Applied Biosystems) using following conditions: 15 cycles of 96°C for 10 s, 55°C for 5 s, 70°C for 1 min and 15 cycles of 96°C for 10 s and 70°C for 1 min. The products were separated on 4.75% polyacrylamide gels using ABI 373 semiautomated sequencer (Applied Biosystem, Warrington, UK) and analysed using Sequencing Analysis version 3.0 and Factura version 2.0.1 (PE Applied Biosystems). PCR reactions were repeated at least once for all homozygous samples to distinguish the true homozygotes from PCR failures (allele drop out).

5.2.2.2 *HLA-DRB1* PCR-SSP genotyping

Nineteen PCR reactions were used to genotype *HLA-DRB1* alleles **01-16* and to subtype the *HLA-DRB1*01* alleles. All the primers used in the reactions are listed in table 5.1. Control primers amplifying a 796 bp fragment from the third intron of *HLA-DRB1* were included in the PCR reactions. PCR amplification was performed in a 15 μl reaction mixture containing 60 ng DNA, 67 mM Tris-HCl, 16 mM $(NH_4)_2SO_4$, 2 mM $MgCl_2$, 200 μM of each dNTP, 3.4 μM of each allele-specific primer, 0.1 μM of each control primers and 0.4 units of Bioline Taq polymerase. The following cycling conditions were used: 96°C for 1 min, 5 cycles of 96°C for 25 s, 70°C for 45 s, 72°C for 45 s, 21 cycles of 96°C for 25 s, 65°C for 50 s, 72°C for 45 s and 4 cycles of 96°C for 25s, 55°C for 1 min, 72°C for 2 min (Bunce et al. 1995). The products were resolved and visualised on ethidium bromide stained agarose gels.

Table 5.1 Primers used in the *HLA-DRB1* genotyping by PCR-SSP.

The lengths of the PCR fragments are given in base pairs.

Alleles recognized	Orientation	Primer sequences 5' - 3'	Product Length
DRB1*0101, 01021, 01022, 0104-6	Sense Antisense	TTGTGGCAGCTTAAGTTTGAAT CCGCCTCTGCTCCAGGAG	194
DRB1*0101, 0103, 0105	Sense Antisense	TTGTGGCAGCTTAAGTTTGAAT CTGCACTGTGAAGCTCTCAC	254
DRB1*01021, 01022, 0104, 0106	Sense Antisense	TTGTGGCAGCTTAAGTTTGAAT CTGCACTGTGAAGCTCTCCA	254
DRB1*0103	Sense Antisense	TTGTGGCAGCTTAAGTTTGAAT CCCGCTCGTCTTCCAGGAT	195
DRB1*1501-9	Sense Antisense	TCCTGTGGCAGCCTAAGAG CCACCGCGGCCCGCGC	206
DRB1*1507, 1601-5, 1607-8	Sense Antisense	TCCTGTGGCAGCCTAAGAG CTCCGTCACCGCCCGGT	136
DRB1*0301-17, 1107	Sense Antisense	GTTTCTTGGAGTACTCTACGTC GTCCACCCGGCCCCGCT	211
DRB1*0401-2, 0404-5, 0408-10, 0413-16, 0419, 0421, 0423, 0424, 0426, 0428-30, 0432-33, 0435-36, 1122	Sense Antisense	GTTTCTTGGAGCAGGTTAAACA CTGCAGTAGGTGTCCACCG	222
DRB1*0402-4, 0406, 0410-13, 0415, 0418, 0422, 0425, 0427, 0432, 0436, 1410	Sense Antisense	GTTTCTTGGAGCAGGTTAAACA CTGCACTGTGAAGCTCTCCA	259
DRB1*0308, 1101-4, 1106-21, 1123-29, 1132, 1137, 1139, 1141	Sense Antisense	GTTTCTTGGAGTACTCTACGTC CTGGCTGTTCCAGTACTCCT	175
DRB1*1201-1202, 12032, 1205, 1206	Sense Antisense	AGTACTCTACGGGTGAGTGTT CTGTTCCAGGACTCGGCGA	162
DRB1*1102-3, 1111, 1114, 1116, 1120-21, 1141, 1301-2, 1304, 1308, 1315-6, 1319-20, 1322-24, 1327-29, 1331-32, 1334-36, 1338-41, 1343, 1345, 1416	Sense Antisense	GTTTCTTGGAGTACTCTACGTC TCCACCGCGGCCCGCTC	210

Alleles recognized	Orientation	Primer sequences 5' - 3'	Product Length
DRB1*1116, 1120, 1301-2, 1308, 1316, 1320, 1327-29, 1331-32, 1334-36, 1339-41, 1343, 1416	Sense Antisense	GTTCCTGGACAGATACTTCC TCCACCGCGGCCCGCTC	152
DRB1*0301-7, 1309, 1311, 1314-17, 0820, 1301-2, 1305-11, 1314-16, 1318-20, 1322-25, 1327-29, 1334-37, 1340-42, 1344, 1402-3, 1406, 1409, 1412, 1414, 1417, 1419-21, 1423-24, 1427, 1429-30, 1433, 1436	Sense Antisense	GTTTCTTGGAGTACTCTACGTC GGCTGTTCCAGTACTCGGCATC	173
DRB1*1107, 1401, 1405, 1407-8, 1414, 1418, 1423, 1426, 1433, 1435-36	Sense Antisense	GTTTCTTGGAGTACTCTACGTC TCTGCAATAGGTGTCCACCT	223
DRB1*0701	Sense Antisense	GGTGCAGTTGCTGGAAAGACT CCCGTAGTTGTGTCTGCACAC	183
DRB1*0801-13, 0815-16, 0818-19, 0822, 1404, 1411, 1415, 1428, 1431	Sense Antisense	AGTACTCTACGGGTGAGTGTT CTCCGTCACCGCCCGGT	130
DRB1*09012	Sense Antisense	GACGGAGCGGGTGCGGTA CCCGTAGTTGTGTCTGCACAC	192
DRB1*10011-12	Sense Antisense	CGGTTGCTGGAAAGACGCG CTGCACTGTGAAGCTCTCAC	203
Control primer forward Control primer reverse	Sense Antisense	TGCCAAGTGGAGCACCCAA GCATCTTGCTCTGTGCAGAT	796

5.2.3 *HLA-B* locus genotyping

Limited *HLA-B* locus genotyping was performed by PCR-SSP in order to assess *HLA-B27* homozygosity and heterozygosity. One PCR reaction was carried out to determine the *HLA-B27* carrier status (Bunce et al. 1995), and five PCR reactions were carried out to determine all the other common *HLA-B* alleles. Primers and alleles that they specify are described in tables 5.2 and 5.3. The control primers used in the reactions are described in table 5.1. PCR amplification was performed as described above (see section 5.2.2.2). Figure 5.1 illustrates positive reactions for *HLA-B27* PCR reaction (a) and *HLA-B27* heterozygote genotyping (b).

Table 5.2 Sequences of the *HLA-B* locus primers used in this study.

Primer mix number	Orientation	Primer sequences 5' - 3'	Location
62	Sense	GCTACGTGGACGACACGCT	HLA-B exon 2
	Antisense	TCTCGGTAAGTCTGTGCCTT	HLA-B exon 2
2336	Sense	ACACAGATCTGCAAGACCAAC	HLA-B exon 2
	Antisense	CCCCAGGTCGCAGCCG	HLA-B exon 3
2337	Sense	ACACAGATCTGCAAGACCAAC	HLA-B exon 2
	Antisense	CCTTGCCGTCGTAGGCGTA	HLA-B exon 3
2339	Sense	ACACAGATCTGCAAGACCAAC	HLA-B exon 2
	Antisense	TGTCCGCCGCGGTCCAG	HLA-B exon 3
1113	Sense	GCGAGGGGACCGCAGGC	HLA-B intron 1
	Antisense	GCGCAGGTTCCGCAGGC	HLA-B exon 2
323	Sense	CGCGAGTCCGAGGATGGC	HLA-B exon 2
	Antisense	CAGGTATCTGCGGAGCCA	HLA-B exon 3

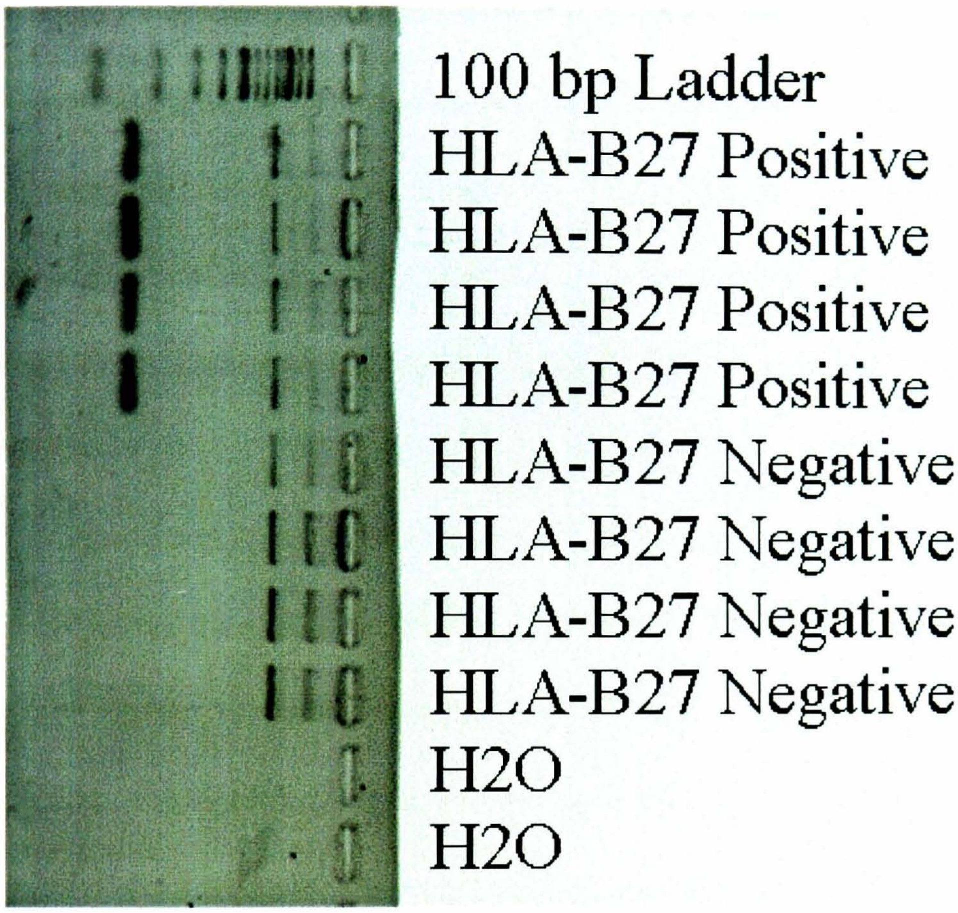
Table 5.3 Alleles recognised by the primers.

Primer mix number	Alleles recognized	PCR fragment length
62	B*2701-9	150
2336	B*0703/8/16/27, *0801-08/10-16, *1310, *1405/6, *1501/3/5/7-12/14/15/18/19/23/24/28-31/33-35/37-40/43/45-58/60/61/63-66/68-73, *1801/3-15/17, *3505/12/16/17/22/30-32/39/43/44, *3801-9, *3901-8/10-20/22-27, *4001-3/5/7-16/18-27/29/31-33/35-40/42/43, *4102-4, *4402-6/8/9/11-14/16/17/19/21-27/29-32, *4801/3-7, *5004, *5106	369
2337	B*0703/8/16/27, *0801-13/15/16, *1509/10/30/37/45/48/63, *3502/4/9/12/18/22/31/34/39/44, *3914, *4001/2/4-16/18/19/21-26/28-35/37/39/40/42-44, *4101-6, *4405/25/31, *4801/3-6, *5101-12/14/16-24/26-32, *5201-5, *7801-5	421
2339	B*0801-5/7/9-16, *1301-4/6-10, *1401-6, *1501-15/18-21/23-40/42-58/60-66/68-73, *1801-15/17/18, *3501-17/19-30/32-44, *3701/3-5, *3801-9, *3901-20/22-27, *4014/26/28, *4101-6, *4402-30/32, *4501-5, *4802, *4901-3, *5001/2/4, *5101-9/11-24/26-32, *5201-5, *5301-9, *5901, *7801-5	469
1113	B*0702-10/12-26/28-31, *0801/4/5/7-16, *1309, *1401-6, *1501-12/14/15/18-21/25-35/37-40/42/44-56/58/60-66/68-73, *1801-8/10-15/17/18, *2708/12/18, *3501-26/28-44, *3901-19/22-27, *4001-12/14-16/18/20-36/38-40/42-44, *4101-4/6, *4201/2/4, *4409, *4501-6, *4601/2, *4702/3, *4801-7, *5001/2/4, *5401/2, *5501-5/7-11, *5601-6/8-10, *6701/2, *7301, *7801-5, *8101, *8201/2, *8301	376
323	B*1301-4/6-10, *1501/2/4-8/11/13/15-17/20/21/24-28/30-36/39/40/42-45/48/50/55-58/60/63/65-67/70/71/73, *4021, *4601/2, *5701-6/8/9	643

Figure 5.1 ***HLA-B27*** **homozygote/heterozygote genotyping.**

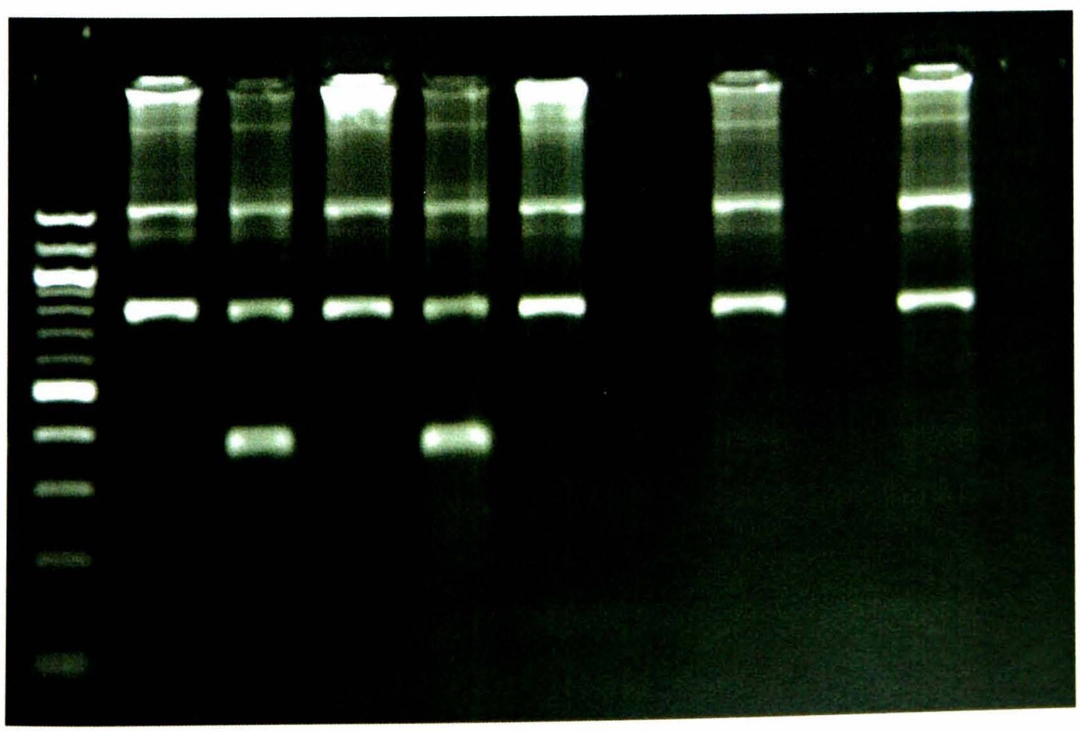
a) *HLA-B27* PCR reaction, b) *HLA-B27* heterozygote genotyping using primer mix number 2336. Lane 1 – 100 bp ladder; Lanes 3 and 5 – positive reactions; Lanes 2, 4, 6, 8 and 10 – negative reactions; Lanes 7 and 9 – H_20

a)



b) Lanes:

1 2 3 4 5 6 7 8 9 10



5.2.4 Statistical Analysis

Haplotype reconstruction. The SIMWALK2 program (version 2.83) was used for haplotype reconstruction in all the AS families (Sobel and Lange 1996). Only families where at least two individuals were fully genotyped for *HLA-DRB1* and *HLA-B27* heterozygote/homozygote were included in the analysis. In order to assess independent effects, only one affected invidual per family was chosen as a case for the case-control analysis. SIMWALK2 provides the most likely solution for haplotype reconstruction, but does not assign a probability that the assignment is correct.

The case-control study. For the case-control study the *HLA-DRB1* results from the sequencing, the primer extension reactions and the PCR-SSP were pooled to correspond to the classic *HLA-DRB1* specificities **01* to **10*, as this was the typing resolution in the control population. Haplotype frequencies in AS cases and the control population were compared by the chi-squared analysis. The relative predispositional effect (RPE) method was used to evaluate the relative effects of *HLA-DRB1* alleles (Payami et al. 1989).

LD and power calculation. The LD between *HLA-B27* and *HLA-DRB1* alleles was calculated using the Lewontin's standardised disequilibrium coefficient D' (Lewontin 1988), calculated using the program 2BY2 (Lathrop and Lalouel 1984). The proband and founder haplotypes estimated using the program SIMWALK2 were used as the input (Sobel and Lange 1996). The statistical significance of the finding was assessed using the chi-squared test. The power of the study was determined by the Genetic Power Calculator (http://statgen.iop.kcl.ac.uk/gpc/) (Purcell et al. 2003).

Disease severity data. QTDT and analysis of covariance (ANCOVA) were used to calculate the significance of differences in continuous variables between different HLA groups. The Kolmogorov-Smirnov test was used to test for deviations from the normal distribution prior to ANOVA analysis using SPSS software (Kinnear and Gray 2000). In case of non-normality of the variable and where stated, logarithmic or reciprocal transformation was applied in order to achieve normality. Reciprocal transformation was employed only if the distribution was considerably more positively skewed than lognormal. BASFI and BASDAI correlated with disease duration and gender in this data set and they were treated as covariates in QTDT and ANCOVA analyses. 2x2 contingency table using chi-squared analysis was performed to analyse the clinical manifestations of the disease. All the p-values shown are 2-tailed, and p-values less than 0.05 were considered significant. All the p-values are uncorrected for multiple comparisons.

The calculation of the expected number of HLA-B27 homozygotes. The antigen frequency of *HLA-B27* was determined from a previously published source of Finnish Bone Marrow Donor Registry based on 10000 samples (Siren et al. 1996). One thousand and four hundred and forty-four of the 10000 individuals were *HLA-B27* positive corresponding to an antigen frequency of 14.4%. The expected frequency of homozygotes for *HLA-B27* in healthy controls was calculated assuming HWE with regard to *HLA-B27*. Let the *HLA-B27* allele frequency be a and non-*HLA-B27* b such that $a+b=1$. The likelihood of being *HLA-B27*-antigen positive is given by:

P (B27/B27 or B27/nonB27)

$= a^2 + 2ab$

$= a^2 + 2a\,(1\text{-}a) = 0.144$

$a^2 + 2ab + b^2 = 1$

$b^2 = 1 - 0.144$

$b = 0.925$

$a = 1\text{-}b = 0.0748$

The probability of being *HLA-B27* homozygote in individuals known to be *HLA-B27*-positive is given by:

P (B27/B27 | B27/B27 or B27/nonB27)

$= a^2 / (a^2 + 2ab)$

$= a^2 / (a^2 + 2a\,(1\text{-}a))$

$= 0.0748^2/0.144$

$= 0.039\ (3.9\%)$

Familial vs. sporadic disease. All the probands and some of the other affected family members answered a structured questionnaire (appendix 1). AS cases, who had one or more first-degree relatives (father, mother, sister, brother, son, daughter) affected by AS, were considered to have familial disease (see questions Q65-Q66 in the appendix 1). Cases were considered to have sporadic AS if they did not have first-degree relatives with AS. Every subject with familial AS were sex and age matched (within 5 years) with 2 control cases with sporadic AS. Clinical characteristics and HLA genes between familial and sporadic AS cases were compared. Qualitative and quantitative data were analysed using the chi-squared test and ANOVA, respectively.

5.3 RESULTS

5.3.1 Demographic characteristics of the data set

In this data set, male:female ratio was 2.1:1. One hundred and twenty-six (48%) subjects also had AAU, 18 (7%) IBD, 6 (2%) psoriasis and 210 (78%) peripheral arthritis. The mean age at symptom onset (± S.D.) was 27 (±9) years, the mean age at diagnosis 35 (±10) years, the mean diagnostic delay 8 (±8) years, the mean disease duration 24 (±11) years, the mean age 51 (±11) years, mean BASDAI 5.2 (±1.9) and mean BASFI 3.9 (±2.3). No Mendelian inconsistencies were noted in the *HLA-B* locus or *HLA-DRB1* locus. Table 5.4 describes the clinical characteristics of the AS cases in this data set.

The Kolmogorov-Smirnov test demonstrated significant left skewing from the normal distribution for the age of symptom onset (p=0.001), the age of symptom onset after logarithm transformation (p=0.04) and the BASFI (p=0.001). The reciprocal transformation of the age of symptom onset, BASDAI and logarithm transformation of the BASFI did not result in significant skewing from the normal distribution, p=0.39, p=0.15 and p=0.57, respectively. These parameters were used in the ANCOVA and ANOVA analysis, respectively.

Table 5.4 Clinical description of the AS cases used in this study.

All cases with AS and cases stratified according to their *HLA-B27* status are presented. Some families included multiple AS cases. P-values refer to the statistically significant difference between *HLA-B27* positive and negative cases with AS.

	ALL CASES (N=270)	*HLA-B27*-POSITIVE CASES (N=251)	*HLA-B27*-NEGATIVE CASES (N=19)	P-VALUE
AAU (%)	126 (48)	123 (50)	3 (16)	0.008
IBD (%)	18 (7)	15 (6)	3 (16)	NS
PSORIASIS (%)	6 (2)	6 (2)	0	NS
PERIPHERAL ARTHRITIS (%)	210 (78)	197 (78)	13 (68)	NS
AVERAGE BASDAI ± S.D.	5.2 ± 1.9	5.2 ± 1.8	5.1 ± 2.4	NS
AVERAGE BASFI ± S.D.	3.9 ± 2.3	3.9 ± 2.3	4.3 ± 2.3	NS
MALES (%) : FEMALES (%)	182 (67) : 88 (33)	169 (67) : 82 (33)	13 (68) : 6 (32)	NS
AVERAGE AGE OF SYMPTOM ONSET ± S.D. (range)	27.4 ± 8.6 (14-55)	26.7 ± 8.0 (14-55)	35.7 ± 11.2 (18-54)	0.0001
AVERAGE AGE AT DIAGNOSIS ± S.D. (range)	35.4 ± 9.9 (18-73)	34.8 ± 9.6 (18-73)	43.4 ± 10.0 (25-55)	0.0001
AVERAGE DIAGNOSTIC DELAY ± S.D. (range)	8.3 ± 8.3 (0-53)	8.4 ± 8.4 (0-53)	6.9 ± 6.8 (0-26)	NS
AVERAGE DISEASE DURATION (range)	23.5 ± 10.5 (3-60)	23.9 ± 10.2 (3-60)	18.3 ± 12.0 (3-49)	NS
AVERAGE AGE ± S.D. (range)	50.8 ± 11.1 (25-80)	50.6 ± 11.1 (25-80)	54.1 ± 10.1 (32-70)	NS

5.3.2 *HLA-B27*

5.3.2.1 The frequency of *HLA-B27* positive haplotypes in cases and controls

HLA-B27 antigen was detected in 251 (93%) of the 270 AS cases. Two hundred and twenty-one (52%) of the 421 haplotypes of AS probands reconstructed by the SIMWALK2 program were *HLA-B27* positive, compared with 105 (8.4%) of the 1254 control haplotypes. This increase was highly significant (χ^2=389, p< 10^{-10}, OR=12, 95% CI=9-16).

5.3.2.2 The expected and observed number of homozygotes for *HLA-B27*

Twenty-seven probands with AS were homozygotes for *HLA-B27*, 211 probands were heterozygotes and 19 probands were *HLA-B27* negative. Assuming HWE with regard to *HLA-B27*, the expected number of *HLA-B27* homozygotes of 257 probands is 9 (0.039 x 238) and the expected number of heterozygotes is 229 (238-9). As shown in table 5.5, there was an over-representation of *HLA-B27* homozygotes in the proband group (χ^2=9.7, p=0.002, OR=3.3, 95% CI=1.6-6.8). Table 5.6 describes the clinical characteristics of *HLA-B27* homozygous and heterozygous AS cases. *HLA-B27* homozygosity was significantly associated with a decrease in BASDAI (mean BASDAI: *HLA-B27* homozygotes 4.5 ± 1.6, *HLA-B27* heterozygotes 5.4 ± 1.8, p=0.003). No other significant associations were noted between clinical characteristics and *HLA-B27* homozygosity.

Table 5.5 The expected and observed number of *HLA-B27* homozygotes of the 238 *HLA-B27* positive probands.

The expected number of homozygotes was calculated assuming HWE with regard to *HLA-B27* (see methods section 5.2.4).

	Observed	Expected
HLA-B27 homozygote	27 (11%)	9 (4%)
HLA-B27 heterozygote	211 (89%)	229 (96%)
Total	238	238

Table 5.6 Clinical characteristics of *HLA-B27* homozygous and heterozygous cases.

	ALL *HLA-B27*-POSITIVE CASES (N=251)	*HLA-B27*-HOMOZYGOUS CASES (N=31)	*HLA-B27*-HETEROZYGOUS CASES (N=220)	P-VALUE
AAU (%)	123 (50)	19 (61)	104 (47)	NS
IBD (%)	15 (6)	0	15 (7)	NS
PSORIASIS (%)	6 (2)	2 (6)	4 (2)	NS
PERIPHERAL ARTHRITIS (%)	197 (78)	21 (68)	176 (80)	NS
AVERAGE BASDAI ± S.D.	5.2 ± 1.8	4.5 ± 1.6	5.4 ± 1.8	0.003
AVERAGE BASFI ± S.D.	3.9 ± 2.3	3.5 ± 2.6	4.0 ± 2.3	NS
AVERAGE AGE OF SYMPTOM ONSET ± S.D. (range)	26.7 ± 8.0 (14-55)	25.9 ± 7.3 (15-44)	26.8 ± 8.1 (14-55)	NS

5.3.2.3 Disease severity and *HLA-B27*

Table 5.4 describes the cases with AS stratified according to their *HLA-B27* status. One hundred and twenty-three (98%) of the 126 AS cases with AAU were *HLA-B27*-positive, which was significantly increased compared with the 122 (88%) of the 138 AS cases without AAU (χ^2=7.05, p=0.008, OR=5.4, 95% CI=1.6-19). No significant differences were noted between AS associated with IBD, psoriasis or peripheral arthritis and the carriage of *HLA-B27* in this data set.

HLA-B27 positive cases had a significantly younger age of symptom onset. This association was noted using both QTDT (*HLA-B27* is associated with younger age of symptom onset by 5.3 years, p=0) and ANOVA (mean age of symptom onset: *HLA-B27*-positive 26.7 years, *HLA-B27*-negative 35.7 years, p=1.0 x 10^{-4}). *HLA-B27* was also associated with younger age at diagnosis by 5.5 years (p=0). A marginal association was noted between BASFI and *HLA-B27* using QTDT (p=0.04), but this was not significant in ANCOVA analysis (p=0.10).

5.3.3 *HLA-DRB1-B27* haplotypes and disease susceptibility and severity

5.3.3.1 Case-control study – disease susceptibility

The frequencies of *HLA-DRB1-B27* haplotypes between AS cases and controls are presented in table 5.7. *HLA-B27*-positive cases and controls were compared. No significant associations were noted.

Table 5.7 The number of *HLA-DRB1-B27* haplotypes in AS cases and controls.

	HLA-B27 positive AS haplotypes	*HLA-B27* positive control haplotypes	OR	p-value
	N (%)	N (%)		
DRB1*01	57 (26)	26 (25)	1.1	0.84
DRB1*02	10 (5)	8 (8)	0.6	0.25
DRB1*03	5 (2)	2 (2)	1.1	0.84
DRB1*04	50 (23)	25 (24)	0.9	0.81
DRB1*05	9 (4)	6 (6)	0.7	0.51
DRB1*06	15 (7)	4 (4)	1.7	0.28
DRB1*07	6 (3)	0	6.4	0.09
DRB1*08	66 (30)	32 (30)	1.0	0.91
DRB1*09	1 (0.5)	2 (2)	0.3	0.20
DRB1*10	2 (1)	0	2.4	0.33
Total	221	105		

5.3.3.2 *HLA-DRB1-B27* haplotypes and disease severity

Using the total evidence of association within the QTDT package, several weak associations were noted between quantitative traits and *HLA-DRB1-B27* haplotypes.

*HLA-DRB1*09-B27* haplotype is associated with -2.8 decrease in BASDAI, χ^2=5, p=0.03. *HLA-DRB1*04-nonB27* is associated with 0.7 increase in BASFI, χ^2=5, p=0.03. *HLA-DRB1*08-nonB27* haplotype was associated with -0.8 decrease in BASFI, χ^2=5, p=0.02. *HLA-DRB1*03-nonB27* haplotype was associated with older age of symptom onset by 6 years, χ^2=8, p=0.006.

5.3.4 *HLA-DRB1* and the clinical manifestations of the disease

*HLA-DRB1*12* is associated with -1.5 decrease in BASFI, χ^2=5, p=0.03. *HLA-DRB1*08* is associated with younger age of symptom onset by 2 years, χ^2=4, p=0.05. *HLA-DRB1*03* is associated with older age of symptom onset by 6 years, χ^2=11, p=0.001. *HLA-DRB1*13* is associated with older age of symptom onset by 2 years, χ^2=4, p=0.05. The disease severity scores did not correlate with the age of symptom onset. No significant associations between *HLA-DRB1* alleles and AS complicated by AAU, IBD or peripheral arthritis were seen.

5.3.5 *HLA-DRB1* sequencing results

One hundred ninety-four alleles from 128 probands were sequenced. One hundred and forty-six alleles (75%) were unambiguously resolved to specify a certain *HLA-DRB1* subtype. Twenty-five alleles (13%) remained partially ambiguous and could specify 2 possible subtypes. Twenty-three alleles (12%) remained ambiguous and could specify 3 or more alleles.

Table 5.8 specifies the *HLA-DRB1* subtypes identified by direct sequencing. Thirteen different subtypes were identified unambiguously. The only *HLA-DRB1*01* subtype identified was *HLA-DRB1*0101*. The predominant subtype of *HLA-DRB1*04* was **0408*, other *HLA-DRB1*04* subtypes identified were **04011, *04031* and **0404. HLA-DRB1*02* subtypes identified were **15011* and **16011*.

The allele frequency of *HLA-DRB1*08* (1% by sequencing) was significantly reduced in comparison with a previous report, in which the *HLA-DRB1*08* allele frequency was 9% (Siren et al. 1996). Additionally, the sequencing of the product of the group specific primer *HLA-DRB1*03/08/11/13/14* repeatedly resulted in ambiguous (94% of all the ambiguities) or failed sequence. Thus, all the samples which remained homozygote after sequencing (n=58) were subsequently genotyped for *HLA-DRB1*03, *08, *11, *13* and **14* using the PCR-SSP approach (Bunce et al. 1995).

Table 5.8 ***HLA-DRB1*** **subtypes identified by direct sequencing in probands.**

Ambiguous sequence is denoted xx, which could specify 3 or more possible subtypes.

Classic HLA-DRB1 type	Subtypes identified	Number of alleles (%)
DRB1*01	0101	55 (28.4)
DRB1*02	15011	23 (11.9)
	16011	3 (1.5)
	15xx	2 (1.0)
DRB1*03	03011	7 (3.6)
	03011 or 0305	6 (3.1)
DRB1*04	04011	17 (8.8)
	04031	2 (1.0)
	0404	7 (3.6)
	0408	20 (10.3)
	0404 or 0423	5 (2.6)
	04xx	4 (2.1)
DRB1*05	11011 or 11012	4 (2.1)
	1103 or 1111	1 (0.5)
	11xx	5 (2.5)
	1201 or 1206	7 (3.6)
DRB1*06	1301	3 (1.6)
	1302	2 (1.0)
	13xx	11 (5.7)
DRB1*07	0701	4 (2.1)
DRB1*08	0801 or 0806	2 (1.0)
DRB1*09	09012	1 (0.5)
DRB1*10	1001	2 (1.0)
	10xx	1 (0.5)
		194

The allele frequencies of *HLA-DRB1*04* subtypes were compared between the current and a historic study (Westman et al. 1994a) (table 5.9). *HLA-DRB1*0408* allele was significantly increased in AS cases of this study compared to the historic cadaveric donors (0.0002). Respectively, the *HLA-DRB1*0401* subtype allele frequency was decreased in AS cases compared to cadaveric donors (p=0.004). The *HLA-DRB1*0404* subtype frequency was significantly decreased in AS cases of the current study and the historic *HLA-B27* positive individuals (p=0.04).

Table 5.9 Comparison of *HLA-DRB1*04* subtype allele frequencies between the current and historic study.

Forty-six alleles were sequenced unambiguously. The partially ambiguous sequences indicating either *HLA-DRB1*0404* or **0423* subtype were included in the *HLA-DRB1*0404* group as the **0423* subtype is not reported to occur in the Finnish population and the exclusion of these ambiguous genotypes would bias the allele frequencies. P-values refer to the significant difference between the AS cases and healthy *HLA-B27* positive individuals or random cadaveric donors.

	Current study	**Historic study**					
HLA-DRB1*04 subtypes	**AS cases (N=51)**	**AS cases (N=14)**	**ReA cases (N=13)**	**HLA-B27+ controls (N=26)**	**p-value**	**Cadaveric donors (N=50)**	**p-value**
	N (%)	N (%)	N (%)	N (%)		N (%)	
***0401**	17 (33)	2 (14)	5 (38)	5 (19)	NS	31 (62)	0.004
*0402	0	0	0	1 (4)	NS	0	NS
*0403	2 (4)	2 (14)	0	1 (4)	NS	3 (6)	NS
***0404**	12 (24)	7 (50)	4 (31)	12 (46)	0.04	11 (22)	NS
*0407	0	0	0	1 (4)	NS	1 (2)	NS
***0408**	20 (39)	3 (21)	4 (31)	6 (23)	NS	4 (8)	0.0002

5.3.6 Comparison of clinical features and the carriage of *HLA-B27* between familial and sporadic cases

Clinical characteristics of the sporadic and familial cases are described in table 5.10. The mean age (± S.D.) in both the familial and sporadic AS cases is 50.5 ± 10.9 years and male:female ratio is 1.3. No significant differences were noted in *HLA-B27* status or clinical characteristics between familial and sporadic disease. The trend is that familial AS cases exhibit a younger age of symptom onset, but this was not statistically significant (mean age of symptom onset: familial AS 25.7 years, sporadic AS 27.7 years, p=0.10). Sixty-one (97%) of the 63 familial AS cases were *HLA-B27* positive compared with the 113 (94%) of the 120 sporadic AS cases. This increase was not statistically significant (χ^2=0.62, p=0.43). Eleven (18%) of the 61 *HLA-B27* positive familial AS cases were *HLA-B27* homozygote compared to the 12 (11%) of the 111 sporadic AS cases. This was not statistically significant (χ^2=1.9, p=0.17).

Table 5.10 Clinical description of familial and sporadic AS cases.

The differences between the familial and sporadic cases were not statistically significant.

	FAMILIAL AS CASES (N=68)	SPORADIC AS CASES (N=136)
AAU (%)	34 (50)	61 (45)
IBD (%)	6 (9)	8 (6)
PSORIASIS (%)	2 (3)	7 (5)
PERIPHERAL ARTHRITIS (%)	51 (75)	104 (76)
AVERAGE BASDAI ± S.D.	5.1 ± 2.0	5.3 ± 1.8
AVERAGE BASFI ± S.D.	4.1 ± 2.6	3.8 ± 2.2
AVERAGE AGE OF SYMPTOM ONSET ± S.D. (range)	25.7 ± 6.8 (11-44)	27.7 ± 8.4 (16-56)
AVERAGE AGE AT DIAGNOSIS ± S.D. (range)	34.3 ± 8.7 (20-54)	35.6 ± 9.5 (18-63)
AVERAGE DIAGNOSTIC DELAY ± S.D. (range)	8.5 ± 6.9 (0-35)	8.0 ± 7.3 (0-29)
HLA-B27 POSITIVE (%)	61 (97) (N=63)	113 (94) (N=120)
HLA-B27 HOMOZYGOTE (%) : HLA-B27 HETEROZYGOTE (%)	11 (18) : 49 (82)	12 (11) : 99 (89)

5.3.7 Power of the study

The power of the case-control study is presented in table 5.11.

Table 5.11 The power of the case-control study using 221 case haplotypes and 105 control haplotypes.

The power is calculated using an AS prevalence of 0.13% and GRR of 2.

	DISEASE MODEL		
	Dominant	Codominant	Recessive
HLA-DRB1*01	72%	47%	16%
HLA-DRB1*02	65%	33%	7%
HLA-DRB1*03	48%	21%	5%
HLA-DRB1*04	72%	46%	15%
HLA-DRB1*05	50%	22%	5%
HLA-DRB1*06	67%	35%	7%
HLA-DRB1*07	31%	13%	5%
HLA-DRB1*08	72%	44%	13%
HLA-DRB1*09	21%	10%	5%
HLA-DRB1*10	14%	8%	5%

5.3.8 LD calculations

In the proband population LD between *HLA-B27* and several *HLA-DRB1* alleles was noted: *HLA-DRB1*01* (D' = 0.19, p = 0.05), *HLA-DRB1*02* (D'=-0.6, p=9 x 10^{-5}), *HLA-DRB1*03* (D' = -0.47, p = 0.03), *HLA-DRB1*08* (D' = 0.53, p = 2 x 10^{-7}), *HLA-DRB1*11* (D' = -0.43, p = 0.04) and *HLA-DRB1*13* (D' = -0.42, p = 0.002) (table 5.12).

Table 5.12 LD values and the corresponding p-values between *HLA-B27* and *HLA-DRB1* alleles calculated using the proband and founder haplotypes.

Lewontin's standardised disequilibrium coefficient D' (p-value) is shown.

	FOUNDERS	PROBANDS
HLA-DRB1	|D'| (p-value)	|D'| (p-value)
*01	0.06 (0.22)	0.19 (0.05)
*02	-0.38 (0.003)	-0.55 (8.7×10^{-5})
*03	-0.55 (0.001)	-0.47 (0.03)
*04	0.07 (0.21)	0.04 (0.69)
*07	0.005 (0.97)	0.47 (0.20)
*08	0.41 (4.1×10^{-12})	0.53 (2.0×10^{-7})
*09	-0.52 (0.10)	-0.52 (0.27)
*10	0.10 (0.74)	0.30 (0.62)
*11	-0.36 (0.10)	-0.43 (0.04)
*12	-0.46 (0.18)	-0.37 (0.25)
*13	-0.40 (0.001)	-0.42 (0.002)
*14	-0.46 (0.44)	1.0 (0.34)

5.4 DISCUSSION

Previous studies have demonstrated several distinct associations between *HLA-DRB1* alleles and AS susceptibility suggesting that the HLA association in AS may be complex and reflect the impact of combinations of alleles carried on a particular haplotype. Alternatively, the disease association may arise from a combination of *HLA-B27* and other HLA loci with an epistatic interaction between these loci. In this study a Finnish population was used to investigate the influence of *HLA-B27* and *HLA-DRB1* genes and haplotypes in AS. Because of the significant LD within the MHC, *HLA-B27* positive case and control haplotypes were used to assess independent effects.

No statistically significant difference between distinct *HLA-DRB1-B27* haplotypes of AS cases that were adequately matched with *HLA-B27* positive control haplotypes was noted. Due to extreme polymorphic nature of the *HLA-DRB1* locus and the modest effects of HLA class II genes on susceptibility to AS, large-scale case-control studies are required to dissect these effects. The haplotypic case-control study presented in this chapter was underpowered. Using the AS prevalence of 0.13% (Kaipiainen-Seppanen et al. 1997) and the GRR of 2, the power of this study was 5-72% depending on the disease model and allele frequencies. A further set of Finnish control samples with *HLA-DRB1-B27* haplotypes will be available to increase the power of the study (Jorma Ilonen, personal communication).

The control haplotypes used in this study were haplotypes of Finnish parents of children with IDDM not transmitted to the diabetic child, nor found in siblings or parents with IDDM. These haplotypes are thought to be representative of the healthy Finnish population as only the non-transmitted haplotypes of unaffected individuals were selected. Theoretically, these haplotype may include protective genes for IDDM as the haplotypes were never transmitted to, nor found in, diabetic individuals within the families. The haplotypes were obtained from a large previous study (Tuomilehto-Wolf et al. 1989); these haplotypes were used as no healthy Finnish DNA was available. However, a control population of healthy individuals would be preferable to the control population used in this study.

Ethnically matched *HLA-B27* negative control haplotypes were not available for analysis. Preliminary data on the English population noted a positive association between *HLA-DRB1*01* allele and AS on the *HLA-B27* negative strand, but not on the *HLA-B27* positive strand (Sims et al. 2004). This suggests that *HLA-DRB1*01* or another gene in LD with it on the *HLA-B27* negative strand are involved in susceptibility to AS, possibly acting in combination with *HLA-B27*.

The study presented in this chapter had 80% power to exclude an association between *HLA-DRB1*01* allele and disease susceptibility with GRR of 2.2-4.4 on the *HLA-B27* positive strand depending on the assumed disease model (dominant, codominant or recessive). This preliminary data set suggests that at least in the Finnish population *HLA-DRB1* alleles on the *HLA-B27* positive strand, compared with healthy *HLA-B27* matched haplotypes, are unlikely to have a major effect on the disease susceptibility. Further studies using a larger sample size and analysing the effect of *HLA-B27* negative haplotypes in the Finnish population is warranted in the future.

Previous studies have reported an association between *HLA-DRB1*01* and AS, but these studies have not matched the cases and controls for *HLA-B27* at the haplotypic level. Haplotype-based approaches have yielded successful results in the HLA region (Jawaheer et al. 2002; Newton et al. 2003), where the complexity of the LD and the density of genes make the detection of the causative variants a challenging task. The LD varies in a haplotype-specific manner in the HLA region; strong and extensive LD is found across both common and rare haplotypes (Ahmad et al. 2003). The study presented in this chapter fully controlled for the association of *HLA-B27* by comparing the different *HLA-DRB1* alleles between *HLA-B27* positive case and control haplotypes. However, the cases and controls were not matched for different *HLA-B27* subtypes. The *HLA-B*2705* is likely to be the most prominent subtype among both cases and controls. However, the possibility exist that different *HLA-B27* subtypes confer differential strength of association with the disease. Each *HLA-B27* subtype participates in a different HLA haplotype as suggested by different LD pattern between MICA alleles and *HLA-B27* subtypes (Tsuchiya et al. 1998). Thus, ideally the HLA fine mapping studies should match the cases and controls for *HLA-B27* subtypes. In the future large case-control studies employing healthy *HLA-B27* control samples and matching the cases and controls completely for *HLA-B27* subtypes may be better able to investigate the effects of non-*HLA-B27* HLA genes on susceptibility to AS.

Previous studies on the influence of *HLA-B27* homozygosity in the development of AS have been contradictory. Significant excess of HLA homozygotes among the AS cases has been reported (Khan et al. 1978), but this has not been confirmed by others (Suarez-Almazor and Russell 1987). The contribution of *HLA-B27* homozygosity is likely to be modest and large samples sizes are required to investigate this issue. In other studies, homozygosity for HLA has been reported to be associated with susceptibility to autoimmune diseases (Skarsvag et al. 1992), common variable immunodeficiency (De La Concha et al. 1999) and an increased difficulty in clearing infections (Pollicino et al. 1996).

In this study, a large family-based cohort was genotyped for *HLA-B27* homozygosity and heterozygosity. No Mendelian inconsistencies were noted in these families. In the PCR-SSP genotyping panel the primer mix number 1113 that was targeted to define the *HLA-B27* heterozygotes also recognised three *HLA-B27* subtypes, *HLA-B2708, *12* and **18*. Thus, the genotyping assay employed could not detect the *HLA-B27* homozygotes who possess *HLA-B2708, *12* or **18* from the *HLA-B27* heterozygotes. As the frequency of these *HLA-B27* subtypes is known to be very rare (<0.5% in the UK population) and no Mendelian inconsistencies were observed, all the individuals who were positive for this reaction and for *HLA-B27* were treated as *HLA-B27* heterozygotes. This is a conservative approach to investigate the potential increase of *HLA-B27* homozygosity in AS, as it potentially decreases the number of *HLA-B27* homozygotes observed.

HLA-B27 homozygosity was significantly increased from the expected under HWE. Possible explanations include: Threshold model of polygenic disease where the presence of increased number of susceptibility alleles increases the likelihood of developing the disease. Other genetic explanation could be partial dominance of *HLA-B27* in determining the phenotype of AS, although this is unlikely as *HLA-B27* homozygosity was not associated with more severe disease. *HLA-B27* homozygotes may have a higher likelihood to develop abnormal characteristics of HLA-B27 such as homodimeric or misfolded structures. In addition, *HLA-B27* homozygotes may express an increased level of HLA-B27 molecules. The latter is supported by the observation of greater expression of HLA-B27 molecules in patients with AS than in healthy controls (Cauli et al. 2002). It is possible that the genotyping method employed failed to detect some *HLA-B* alleles, which could subsequently increase the number of apparent homozygotes. This could be overcome by performing full *HLA-B* locus genotyping for all the homozygous cases. However, it was estimated that the frequency of the alleles that are not detected by this method is very low. This is supported by the fact that no Mendelian inconsistencies were detected in these families.

A previous report has suggested that *HLA-B27* homozygosity may influence the disease severity (Suarez-Almazor and Russell 1987). A higher frequency of involvement of peripheral joints has been reported among the *HLA-B27* homozygotes (Khan et al. 1978). Surprisingly, this study noted a significant decrease in BASDAI among the *HLA-B27* homozygotes. No other disease severity indices showed significant differences, suggesting that *HLA-B27* homozygote patients may exhibit a milder disease than *HLA-B27* heterozygotes. This is consistent with the whole-genome disease severity screen, in which no linkage between the disease severity indices and the MHC region was observed (Brown et al. 2003).

It has been proposed that *HLA-B27* positive and *HLA-B27* negative AS represent a heterogenous group of phenotypically similar diseases that may have different aetiopathogenic mechanisms. *HLA-B27* positive AS is reported to be associated with younger disease onset (Saraux et al. 1995; Feldtkeller et al. 2003). A recent report suggested that *HLA-B27* is also associated with earlier onset of psoriatic arthritis

(Queiro et al. 2003). The current study showed a significant association between *HLA-B27* and a younger age of symptom onset and a younger age at diagnosis in AS patients confirming the previously reported associations.

An association between *HLA-DRB1*08* and juvenile AS has been reported in Norwegian and Mexican cases (Ploski et al. 1995; Maksymowych et al. 1997b). A weak association between *HLA-DRB1*08* and early age at symptom onset is reported here (section 5.3.4), supporting these previous findings. These findings strongly suggest that both *HLA-B27* and *HLA-DRB1*08*, either independently or as a haplotype, contribute to the genetic susceptibility of early onset disease. Significant LD between *HLA-B27* and *HLA-DRB1*08* was observed, and the association between *HLA-DRB1*08* and younger age of symptom onset may be secondary due to strong LD. However, in this data set *HLA-DRB1*08-B27* haplotype was not associated with the younger age of symptom onset. Also, the association between *HLA-DRB1*08* and age of symptom onset remained significant even when *HLA-B27* was used as a covariate in the analysis, suggesting that the association is likely to be independent. Large studies of *HLA-B27*-negative cases may be better able to dissect whether *HLA-DRB1*08* association is independent, although *HLA-B27* negative AS may be too different to confirm the finding. In contrast, neither *HLA-DRB1*08* nor *HLA-B27* were found to be associated with any of the disease severity scores. Younger age of symptom onset did not correlate with BASDAI or BASFI, suggesting that even if disease occurs earlier in the presence of predisposing genes, it is not more severe. This is consistent with previous studies, indicating that although cases with juvenile onset of AS are more likely to develop hip arthritis, their disease is otherwise similar in character to cases with adult onset disease AS (Calin and Elswood 1988). *HLA-DRB1*03* allele and *HLA-DRB1*03-nonB27* haplotype were associated with older age of symptom onset, suggesting *HLA-DRB1*03*, or another gene on the *HLA-B27* negative strand, may be involved in determining a later age of symptom onset. Overall, these findings suggest that *HLA-DRB1* alleles may influence the age of symptom onset of AS. However, the reported p-values are not corrected for multiple comparisons made, and the associations reported here may merely be attributable to stochastic statistical fluctuations.

AAU is more common in *HLA-B27*-positive than *HLA-B27*-negative cases with AS (Khan et al. 1977; Saari et al. 1978), and this is confirmed by this study. There is no consensus amongst previous studies concerning the association between *HLA-DRB1*08* and AAU. An association has been reported in a Japanese population (Monowarul Islam et al. 1995), and negative findings have been reported in Norwegian and Mexican studies (Ploski et al. 1995; Maksymowych et al. 1997b). A possible explanation is that the juvenile AS may increase AAU. However, a recent study noted that the prevalence of iritis correlates positively with disease duration, but not with age of symptom onset (Brophy and Calin 2001). The present study noted no significant association between *HLA-DRB1*08* and AAU, and thus do not support the Japanese finding. The discrepant associations may be due to different ethnic background of the populations.

Although disease severity has been shown to be primarily genetically determined (Hamersma et al. 2001), there is little information on the role of candidate genes in the severity of AS. As diverse associations between *HLA-DRB1* alleles and the disease phenotype have been reported, it is possible that the discrepant findings may be a consequence of an underlying association with disease severity, rather than susceptibility. In the present study weakly significant associations were observed with

HLA-DRB1 alleles and *HLA-DRB1-B27* haplotypes and BASDAI and BASFI. In a British study, no association was found between *HLA-DRB1*, *HLA-B27* or *HLA-B60* and disease severity indices (Brown et al. 1998a). In a Canadian study no association between *LMP2* polymorphisms and the disease severity scores was observed (Maksymowych et al. 1997a). The whole-genome linkage screen observed no linkage with the disease severity indices and the HLA region, consistent with the general observation that little overlap was noted between regions linked with susceptibility to AS and those linked with the clinical characteristics (Brown et al. 2003). This highlights the difference between these measures. Overall, these findings do not suggest that MHC genes have a major role in determining the severity of AS.

In the healthy Finnish population *HLA-B27* has previously been reported to be in LD with *HLA-DRB1*01*, **04* and **08* (Westman et al. 1994a). Positive LD between *HLA-B27* and *HLA-DRB1*01* and **08*, and negative LD between *HLA-B27* and *HLA-DRB1*02, *03, *11* and **13* was observed. In comparison with the British population the overall LD pattern is similar in the Finnish population (Brown et al. 1998a). In contrast to a previous Finnish report (Westman et al. 1994a), this study did not demonstrate a significant LD between *HLA-DRB1*04* and *HLA-B27*. The sample size in the previous report was relatively small, which may explain the discrepancy. The difference may also be attributable to the methodology used to determine haplotypes and to quantify the extent of LD.

Comparison of *HLA-DRB1*04* subtypes demonstrated that the *HLA-DRB1*0401* and **0408* frequencies were significantly different between AS cases and historic random controls (Westman et al. 1994a). This is consistent with previous reports and confirms the finding that *HLA-B27* associates with *HLA-DRB1*0408* (Westman et al. 1994a). In a previous report *HLA-DRB1*0404* frequency was also increased in *HLA-B27* positive cases compared to random controls (Westman et al. 1994a). This study showed no difference in *HLA-DRB1*0404* frequency between the AS cases and historic random controls. These discrepant results may be attributable to the small sample sizes or bias in the sequencing results due to the difficulty in specifying this subtype from other subtypes.

The clinical characteristics and the *HLA-B27* status were compared between the familial and sporadic AS cases. No significant differences were noted in *HLA-B27* status, BASDAI, BASFI, age of onset, age at diagnosis or prevalence of peripheral arthritis and AAU between familial and sporadic disease. This is consistent with a Dutch report (van der Paardt et al. 2002b). One previous study suggested that familial AS is significantly milder than sporadic disease as assessed by the arthritis impact measurement scales (AIMS) (Calin et al. 1993). This study was not able to confirm this previous finding using BASDAI and BASFI as clinical disease severity indices. The difference may be related to the disease severity indices used, the power of the study or selection of the AS cases. Due to small sample size several AS cases were selected from the same families for the familial group. These AS cases are thus not independent and could bias the analysis.

One of the limitations of this study is the lack of markers genotyped, which hinders the reliable reconstruction of the haplotypes and the definition of the causative variants. As the MHC is completely sequenced and polymorphisms within it are abundant (1999), it may be feasible to employ a dense map of SNPs to determine the primary associations.

The extent of LD in the human MHC region is longer in physical distance (but not in genetic distance) than elsewhere in the genome making it a challenging task to identify the primary associations (Walsh et al. 2003). A haplotype map of the human MHC may facilitate this endeavour and make the mapping of the causative variants quicker and more cost-effective in the future (Walsh et al. 2003).

5.5 CONCLUSIONS

This study supports the concept that the genetic basis of AS is heterogenous, and genes within the human MHC are involved in determining both the disease susceptibility and the age of disease onset. Homozygosity for *HLA-B27* was noted to be significantly increased among the cases with AS. A significant association was noted between age of symptom onset and both *HLA-B27* and *HLA-DRB1*08*. No statistically significant difference between distinct *HLA-DRB1-B27* haplotypes of AS cases that were adequately matched with *HLA-B27* positive control haplotypes was noted. Further studies assessing the effect of *HLA-B27* negative haplotypes and using a larger number of control haplotypes may be better able to clarify the role of *HLA-DRB1* alleles in AS susceptibility.

CHAPTER 6 - THE EFFECT OF POLYMORPHISMS OF THE CYP2D6 GENE IN AS IN A FINNISH POPULATION

This chapter presents the results of a study designed to investigate whether the previously observed association between the Cytochrome P450 2D6 (*CYP2D6*, debrisoquine 4-hydroxylase, OMIM reference number 124030) gene and AS can be demonstrated in a Finnish population. A within-family association study of the main poor metaboliser (PM) allele of the *CYP2D6* gene (CYP2D6*4, OMIM reference number 124030.0001) was performed.

6.1 INTRODUCTION

The association between homozygosity for inactive *CYP2D6* alleles and AS has been reported by two previous independent studies (Beyeler et al. 1996; Brown et al. 2000a). The *CYP2D6* gene was the first non-MHC locus that was associated with AS. The finding was supported by within-family association methods demonstrating a moderate over-transmission of *CYP2D6*4* alleles in patients with AS (p=0.01) and by linkage studies showing a LOD score of 1 (Laval et al. 2001). The strength of association by case-control methods was strong (p=0.0007), but the relative risk of 2.1 only moderate and hence the influence of this gene on overall risk appears small. It has been hypothesized that altered metabolism of a natural toxin or antigen by the CYP2D6 gene may increase susceptibility to AS (Brown et al. 2000a). The aim of this study was to evaluate the effect of *CYP2D6*4* allele in 170 Finnish families with AS.

6.1.1 Cytochrome P450 genes

Cytochrome P450 (CYP) enzymes are a superfamily of microsomal metabolising enzymes. They are involved in the oxidative metabolism of both endogenous and exogenous molecules. They represent the first line of defence against environmental chemicals and are present in all higher organisms. Physiological substrates include steroids, fatty acids, prostaglandins, leukotrienes and biogenic amines. Exogenous compounds include plant toxins, drugs and carcinogens (Johansson et al. 1993). Metabolism of drugs usually converts them to metabolites that are more water soluble and thus more easily excreted by the kidney. It can also convert prodrugs into therapeutically active compounds, and it may even give rise to the formation of toxic metabolites (Weinshilboum 2003). CYP enzymes are the most important of the enzymes that catalyse phase I metabolism, which involves the activation of drugs and carcinogens into more toxic electrophilic intermediates. Phase II drug-metabolising enzymes (such as N-acetyltransferases and glutathione S-transferases) conjugate the intermediates to water-soluble derivatives to complete the detoxification cycle. Approximately 40% of human CYP-dependent drug metabolism is carried out by polymorphic enzymes, which can result in abolished, quantitatively or qualitatively altered, or enhanced metabolism (Ingelman-Sundberg et al. 1999).

6.1.2 CYP2D6 polymorphisms

The *CYP2D6* gene is part of a cluster of three CYP2D genes on 22q13.1. A definite pseudogene *CYP2D8P* and a presumed pseudogene *CYP2D7* are located 5' of the *CYP2D6* locus (Kimura et al. 1989). To date, 77 allelic variants of *CYP2D6* have been identified and characterised (see www.imm.ki.se/cypalleles). Defective alleles can be the result of gene deletions, gene conversions with related pseudogenes and single base mutations causing frameshift, missense, nonsense or splice-site mutations. Many of these polymorphisms are functionally significant, often resulting in altered activity or complete absence of enzyme.

About 5-10% of Caucasians are so-called poor metabolisers (PM) of debrisoquine, completely lacking CYP2D6 activity because of inactivating mutations in both alleles of the *CYP2D6* gene. The PM phenotype causes an impaired ability to metabolise the antihypertensive probe drug debrisoquine and other drugs specific for this drug-metabolising enzyme (Ingelman-Sundberg et al. 1999). Individuals showing impaired activity (intermediate metabolisers) and subjects showing particularly high levels of activity (ultrarapid metabolisers) have been described. The presence of multiple copies of the *CYP2D6* gene has been demonstrated to associate with the ultrarapid metaboliser phenotype (Johansson et al. 1993). Subjects with wild-type *CYP2D6* genotype and the normal range of activity are called extensive metabolisers.

Five PM-associated alleles (*CYP2D6*3, *4, *6, *7, *8*) and *CYP2D6* deletion allele **5* are responsible for the PM phenotype in approximately 99% of Caucasian individuals (Stuven et al. 1996). The most common PM mutation is *CYP2D6*4* allele that contains a splice site mutation ENST00000320841: c.508-3G>A transition at the junction of the intron 4/exon 5. This mutation shifts the position of the 3-prime splice site and results in loss of the first base (G) in exon 5 (Gough et al. 1990). The mutation leads to a stop codon 40 bases further on, giving a truncated transcript and causing absence of the CYP2D6 protein (Kagimoto et al. 1990). The most likely process whereby the transcript is degraded is a form of RNA surveillance known as nonsense-mediated mRNA decay. *CYP2D6*4* mutation represents more than 75% of PM alleles (Broly et al. 1991; Daly et al. 1991). The second most common PM-associated mutation is the gene deletion, *CYP2D6*5*, accounting for 14% of the inactive mutations (Broly et al. 1991).

Medical interest in *CYP2D6* polymorphisms arises because CYP2D6 is responsible for oxidative metabolism of over 30 clinically used drugs and environmental chemicals, including neuroleptics, tricyclic antidepressants, antiarrhythmics and morphine derivatives. Codeine has been demonstrated to be ineffective analgesic in poor metabolisers as morphine is not activated by CYP2D6 (Poulsen et al. 1996). Phenotyping methods use marker substrates specific for CYP2D6 enzyme, such as debrisoquine, metoprolol, dextromethorphan and sparteine, to determine the metabolic status of subjects.

The distribution of variant alleles for *CYP2D6* gene differs markedly between ethnic groups (Table 6.2). *CYP2D6*4* frequencies are approximately 4% in black populations and < 1% in Asians (Nebert 1997). A higher frequency of duplicated or multiduplicated *CYP2D6* genes, indicative of ultrarapid metabolism, has been observed in the Ethiopian population compared to Caucasians (Aklillu et al. 1996).

The origin of the high degree of interspecies and intraspecies variability of drug metabolising enzymes is unknown. It may be the result of continuous molecularly driven coevolution of plants producing phytoalexins and animals responding with new version of enzymes to detoxify these chemicals (Gonzalez and Nebert 1990). The carriers of the polymorphisms may also have a selective advantage such as improved rates of pre-or postnatal development or growth, or resistance to bacterial or viral infections (Nebert 1997).

Genetic differences in the metabolic activation of endogenous compounds or of xenobiotics such as chemical carcinogens have been proposed to increase susceptibility to cancer and other diseases (Lemos et al. 1999). Associations between *CYP2D6* polymorphisms and disease have been extensively studied. Conflicting associations have been reported between *CYP2D6* polymorphisms and systemic lupus erythematosus, lung cancer and Parkinson's disease (Baer et al. 1986; Christensen et al. 1998; Rostami-Hodjegan et al. 1998; Kortunay et al. 1999). The power of most of these association studies has not been reported.

6.2 SUBJECTS AND METHODS

6.2.1 Subjects

At the time of conducting this study 437 individuals from 170 Finnish families were available. The structure of families is in common with chapter three.

6.2.2. Genotyping

The genotyping of *CYP2D6*4* polymorphism was performed using PCR-RFLP as previously described (Brown et al. 2000a). Briefly, a 421 bp fragment was amplified using the primers 5' –GGTGTTCCTCGCGCGCTATG -3' and 5' – CTCGGTCTCTCGCTCCGCAC -3'. A restriction digest was performed using the enzyme BstNI. This enzyme cuts at a polymorphic site at position 1934 in wild type but not mutant alleles. To distinguish between reactions where the restriction digest has failed and homozygous mutant samples, a non-polymorphic BstNI restriction site at position 1772 was also included in a PCR fragment of 421 bp. Thus *CYP2D6*4* homozygotes produce two fragments of lengths 77 and 344 bp, whereas non-*CYP2D6*4* homozygotes produce three fragments of lengths 77, 161 and 183 bp. Amplifications were carried out in 10 µl reactions consisting of 50 ng DNA, 25 ng each primer, 200 µM each dNTP, 2.0 mM $MgCl_2$, 1 µl 10xNH_4 buffer (Bioline, London, UK) and 0.2 units Taq polymerase (Bioline, London, UK). The cycling conditions used were 32 cycles of denaturation at 94°C for 1 min, annealing 60°C for 1 min and extension 72°C for 1 min. Digests, in a total volume of 20 µl, contained 10 µl of PCR product, 2 µl of NE buffer 2 (New England Biolabs, Beverly, MA, USA), 0.2 µl bovine serum albumin, 10 units of BstN1 and dH_20 were incubated for at least 16 hours at 60°C.

6.2.3 Statistics

Within-family association was assessed using a generalisation of the TDT (Clayton 1999). Power calculations were performed as described (Chen and Deng 2001).

6.3 RESULTS

6.3.1 Allele frequencies

Genotype and allele frequencies were in HWE. The founder allele frequency of *CYP2D6*4* was 12% in Finns, shown in table 6.1 with allele frequencies published for other ethnic groups (Ingelman-Sundberg et al. 1999).

Table 6.1 Major variant alleles of the CYP2D6 gene.

The figure is modified after (Ingelman-Sundberg et al. 1999).

CYP2D6 variant	Mutation	Enzyme function	Metabolism	Allele frequencies (%)				
				Finns (this study)	Other Caucasians	Asians	Africans	Ethiopians and Saudi Arabians
*4	Defective splicing	Inactive enzyme	None	12	12-21	1	2	1-4
*5	Gene deletion	No enzyme	None	ND	2-7	6	4	1-3
*2xN	Gene duplication or multiduplication	Higher enzyme levels	Increased	ND	1-5	0-2	2	10-16
*10	Pro34Ser, Ser486Thr	Unstable enzyme	Reduced	ND	1-2	51	6	3-9
*17	Thr107Ile, Arg296Cys, Ser486Thr	Altered substrate affinity	Reduced	ND	0	ND	34	3-9

ND, not determined

6.3.2 *CYP2D6*4* in disease susceptibility

The results of TDT analysis using the program TRANSMIT are presented in table 6.2. No differences between transmitted and untransmitted allele frequencies were found.

Table 6.2 *CYP2D6 1934 G/A* allele untransmitted founder and proband allele frequencies.

		Untransmitted founder		Probands		
Allele		N	%	N	%	P-value
1	G	352	89	351	89	0.69
2	A	44	11	45	11	

6.3.3 Power calculations

Assuming no missing parental genotypes, a population prevalence of AS of 0.1% and recessive disease model, this study had 80% power at the p-value of 0.05 to detect an effect of GRR of 7. Using these same parameters and assuming a locus specific effect of GRR of 2, this study has 9% power, and for 80% power 4405 families would be required.

6.4 DISCUSSION

No association between *CYP2D6*4* alleles and AS could be demonstrated. Replication of genetic data is essential in order to reliably confirm a true association. Failure to replicate a previously reported association may be the result of lack of power, differences in LD patterns between different populations or a false positive initial finding. Previous reports observed an association between homozygosity for *CYP2D6*4* and AS, but not between heterozygosity for *CYP2D6*4* indicating that this is a recessive locus in AS or in LD with another recessive locus (Brown et al. 2000a). TDT has poor power to detect an association for a recessive trait. The low prevalence of PM genotype in a Finnish population further reduces statistical power. This study had only 9% power to detect an association assuming a locus specific GRR of 2. Our power calculations were performed using the TDT power calculator, which allows no missing parents in the calculations. In our case, 58 families had no parents available and 39 families had only 1 parent available. Therefore, the calculated power is an overestimate of the actual power. The study cannot exclude a true association of small locus-specific effect. A larger patient group is required in order to draw definite conclusions.

It is strongly recommended that all the genetic studies report not only the type I error, but also the type II error of the study, that is the estimated power to detect the genetic effect being investigated (Dahlman et al. 2002). Due to positive publication bias and large variance in test statistics, estimates of the size of the locus-specific effect in the initial study tend to be grossly inflated. Thus, the power calculations for a replication study cannot simply be based on the first estimate of the locus-specific effect (Goring et al. 2001; Dahlman et al. 2002).

Large sample sizes are required to have adequate power to identify the non-B27 genetic effects being studied in AS. Using the allele frequency of 0.12, a population prevalence of AS of 0.1% and a recessive disease model, the sample size required to have 80% power at a p-value of 0.05 would be 4405 parent-case trios to detect an effect size of GRR of 2. The recruitment of a patient cohort of this magnitude for a relatively rare disease such as AS is beyond the scope of most centres. Genotyping 4405 parent-case trios is a major undertaking. It has been demonstrated that small changes in susceptibility allele frequencies may have large effects on power (Ardlie et al. 2002b). In case of unknown allele frequencies in the locus of interest, it may be worthwhile to genotype a small subset of samples first and perform power calculations on the basis of these frequencies prior to a further study.

A within-family association method was used to detect a possible association. This has the advantage that the cases and controls are readily matched and spurious associations caused by population stratification are avoided. The disadvantage of TDT test is that it only counts heterozygous parents in the analysis. If the disease allele is rare, as is the case in this study, it reduces the number of heterozygous parents with the mutant allele. TDT requires the genotyping of three individuals (parent-case trios) in each unit compared to two individuals per unit in a case-control study. The recruitment of individuals for a well-matched, large case-control study may be more feasible than a large number of parent-case trios. Alternatively, very large family collections are required, and perhaps international collaborations may facilitate their recruitment.

The genotyping method used for this study does not differentiate the *CYP2D6*4* homozygotes from the *CYP2D6*4/5* heterozygotes. This differentiation would require a further PCR reaction (Brown et al. 2000a). As both *CYP2D6*4* and *CYP2D6*5* alleles are associated with the PM phenotype and the allele frequency of *CYP2D6*4* in the Finnish population is low, *CYP2D6*4* homozygotes and the *CYP2D6*4/5* heterozygotes were not differentiated in this study.

The founder allele frequency of *CYP2D6*4* was 12%. This is below the level for the British (19%) (Brown et al. 2000a), but considerably higher than that in Asians and Africans (1-4%) (Ingelman-Sundberg et al. 1999). The relatively low allele frequency in the Finns compared to the British population may be due to inward gene flow into the Finnish community from the east (Kittles et al. 1998).

It may be worthwhile to study other genes encoding metabolising enzymes in AS. A recent German study demonstrated an association between *CYP1A1*1A* allele and psoriasis, a disease that is clinically related to AS (Richter-Hintz et al. 2003). This allele is associated with lower oxidation of xenobiotics, possibly leading to accumulation of xenobiotics and conceivably a higher chance of aberrant antigen formation. The same study also noted that heterozygosity for *CYP2C19*1A*2A* was protective for psoriatic arthritis, but the sample size for this sub-analysis was very small (Richter-Hintz et al. 2003).

The *CYP2D6* genetic polymorphisms are responsible for pronounced interindividual differences in plasma concentrations and in drug action and side-effects after administration of the same dose. *CYP2D6* poor metabolisers lack the catalytic function of this enzyme and, hence, the plasma concentrations of drugs specific for this enzyme may lead to toxic plasma concentrations and side effects. It has been suggested that genotyping could be used in the clinical setting to customise drug dose (Ingelman-Sundberg et al. 1999). However, most drugs are metabolised by several different enzymes transported by other proteins and eventually interact with one or more targets. Therefore, even if inheritance influences the effect of a drug, the relatively simple one-to-one relation observed for *CYP2D6* may not be obvious. This may explain the relatively few examples of clinical tests based on pharmacogenetics (Weinshilboum 2003).

The environmental trigger in AS is unknown, but is likely to be ubiquitous. Theoretically xenobiotics could elicit or exacerbate AS. Drugs and other chemicals are capable to induce allergic or autoimmune reactions (Griem et al. 1998). One possible mechanism is that self-peptides or MHC determinants are chemically modified. Some metals are believed to induce autoimmune reactions secondary to the creation of new high affinity sites for MHC determinants on self peptides. Alternatively, hapten-specific T lymphocytes may help B lymphocytes that are able to recognise haptenated self-protein (Kimber and Dearman 2002). Examples of autoimmune diseases caused by xenobiotics include Stevens-Johnson syndrome and dihydralazine-induced systemic lupus erythematosus (Brown et al. 2000a). These diseases are clearly different diseases to AS, but demonstrate that impaired xenobiotic metabolism can induce autoimmune disease in humans.

It has been proposed that altered metabolism of a natural toxin by the *CYP2D6* gene could increase susceptibility to AS (Brown et al. 2000a). This hypothesis illustrates how the interaction between an environmental trigger and inter-individual genetic predisposition may confer susceptibility to disease. It is of note that the metabolism of xenobiotics may involve several enzymes and interactions with other proteins. Xenobiotics may interact directly with certain cytokines altering their activity (Brooks et al. 2001), illustrating the potentially complex network of interactions involved in their actions and metabolism.

6.5 CONCLUSIONS

This study failed to demonstrate a significant association between *CYP2D6*4* polymorphism and AS, but the low power of the study cannot exclude the presence of a true positive association. As exemplified by this study large sample sizes are required to detect small effects. The underlying disease model and the disease associated allele frequency are critical factors influencing the test power. Power calculations are essential for the correct interpretation of negative results. The abundance of available information on the existence and function of polymorphisms and the development of high-throughput genotyping means, allow the design of large-scale studies having sufficient power to detect at least common dominant or codominant disease loci. Such studies on different genes for xenobiotics metabolising enzymes could possibly lead to new insights into the AS aetiology in the future.

CHAPTER 7 - DISCUSSION AND FUTURE DIRECTIONS

7.1 BACKGROUND

AS is a complex genetic disease in which allele and locus heterogeneity play important roles. *HLA-B27* is necessary but not sufficient to explain the total variance of the heritability of the disease; other genes are likely to significantly contribute to the disease pathogenesis. The maximization of sample homogeneity is particularly important for the successful outcome of linkage and association studies in AS. This thesis investigates genetic determinants of AS mainly in a homogenous Finnish population.

The current Finnish population of approximately 5 million arose from a limited number of founders about 2000-2500 years ago (de la Chapelle 1993). Since this initial founding, the population has undergone several bottlenecks and genetic drift has shaped the allele frequencies. The Finnish population has proven to be very useful in the mapping of susceptibility loci of rare monogenic disorders (Hastbacka et al. 1994). It has been demonstrated that several disease mutations characterised in Finland show a founder effect. Internal population isolates in the northern Finland reveal greater LD than the older populations (Laan and Paabo 1997; Varilo et al. 2003). In contrast, the general Finnish population does not exhibit larger regions of LD than more outbred populations (Eaves et al. 2000; Taillon-Miller et al. 2000), suggesting that LD mapping projects employing the Finnish population could not reduce the required marker density. The Finnish population may exhibit a limited number of disease genes in initial founders and less genetic heterogeneity. The population is especially suitable for case-control studies because the known and homogenous genetic background of all individuals reduces the effect of population stratification. Furthermore, isolated populations may be more homogeneous in the environment which has the potential to decrease the genetic complexity underlying polygenic diseases. Some linkage-based projects aimed at identifying susceptibility genes in common diseases have been conducted using the Finnish population (Asumalahti et al. 2000; Watanabe et al. 2000; Laitinen et al. 2001). In the light of limited experimental evidence, it remains to be determined whether Finland is more suitable than more outbred population for disease gene mapping of complex diseases.

Significant differences in the allele and haplotype frequencies between the Finnish population and the British population were noted in the *CYP2D6* locus, *TGFB1* locus and HLA region suggesting divergent population histories of these two populations. The extent of observed LD in these regions was not substantially greater in the Finnish population than in the English population, consistent with previous reports (Eaves et al. 2000; Taillon-Miller et al. 2000).

This thesis investigates three candidate regions in AS: *TGFB1* locus on chromosome 19, *CYP2D6* locus on chromosome 22 and the HLA region on chromosome 6. Previously all of these regions have been linked with AS by linkage studies (Rubin et al. 1994; Brown et al. 1998b; Laval et al. 2001).

7.2 SUMMARY OF THE RESULTS

The first project investigates a positional and functional candidate gene, *TGFB1*, in susceptibility to, and severity of, AS in a Finnish and British population. A weak association between the *TGFB1* 1627 C>T polymorphism and AS was noted in the Finnish population (p=0.04), but this was not significant in the English population alone (p=0.34). A marginal association between the *TGFB1* promoter SNP (-800-509) haplotype 2/1 was noted in the English population (p=0.05), but this was not significant in the Finnish data set (p=0.53). These discrepant associations may reflect genuine locus and allele heterogeneity between the Finnish and English populations, different LD patterns with a further truly associated variant, phenotype differences between the populations or different environmental backgrounds. Alternatively, the differences may be related to the stochastic statistical fluctuations and the multiple comparisons made. No association was noted among the families which showed a positive NPL score at this locus, suggesting that these polymorphisms do not explain the linkage of AS susceptibility to chromosome 19.

The second experimental chapter of this thesis describes a novel *HLA-DRB1* genotyping method based on the primer extension reactions. This method was developed to allow rapid and cost-effective screening of a large number of AS cases and their family members. The primary limitation of this method is its low resolution, which could be improved by employing a group of *HLA-DRB1* specific PCR primers.

The third experimental chapter of this thesis describes the effect of HLA genes and haplotypes in susceptibility to, and severity of, AS in the Finnish population. *HLA-B27* positive case and control haplotypes were employed to adequately control for the strong association between *HLA-B27* and AS, and a significant LD between *HLA-B27* and *HLA-DRB1* alleles. As demonstrated in table 5.12, several distinct *HLA-DRB1* alleles exhibit significant LD with *HLA-B27*, and thus it is essential to use *HLA-B27* matched controls. Previous studies have reported positive associations with *HLA-DRB1*01* allele and AS (Brown et al. 1998a; Vargas-Alarcon et al. 2002b), but these studies have not controlled for the effect of *HLA-B27* at the haplotypic level. No statistically significant difference between distinct *HLA-DRB1-B27* haplotypes of AS cases that were adequately matched with *HLA-B27* positive control haplotypes was noted. A further control group will be available for future analysis to increase the power of the study (Jorma Ilonen, personal communication). Homozygosity for *HLA-B27* was over-represented in cases with AS when compared to the expected number of *HLA-B27* homozygotes predicted by Hardy-Weinberg. Significant associations between both *HLA-B27* and *HLA-DRB1*08* and a younger age of symptom onset was found suggesting that genes within the MHC are involved in determining the age of symptom onset.

The fourth experimental chapter of this thesis investigates the influence of *CYP2D6*4* poor metaboliser allele in AS. Due to the low allele frequency of the poor metaboliser alleles in the Finnish population, the power of the within-family association by TDT was poor. No association was found, but due to lack of power this study could not exclude true positive association with the disease either.

7.3 FUTURE DIRECTIONS IN THE GENETICS OF COMPLEX DISEASES

The scope of complex disease gene mapping has been markedly changed by the vast increase in the number of known and putative new genes resulting from the Human Genome Project (Lander et al. 2001), the identification of many polymorphisms (Sachidanandam et al. 2001) and the development of new high-throughput methods for large-scale genotyping (Syvanen 2001). It has also set a new challenge of how best to use the wealth of information. These developments promise great advances in medical genetics. As a result of weak linkage signals, positional cloning has limited use in the identification of complex disease variants conferring low relative risks for individual loci. Attention has therefore turned to association or LD studies in the expectation that this approach may be more effective (Risch and Merikangas 1996).

In theory, association studies can be conducted genome-wide. However no consensus has emerged yet on the number of SNPs that may be required, the selection of study design or the methodology of statistical analysis. Even low-end estimates of the number of SNPs required for a whole genome association study are in the range of 50000 – 100000 (Collins et al. 1999; Boehnke 2000). Two strategies have been proposed for genome-wide association studies namely map-based gene discovery and sequence-based gene discovery (Peltonen and McKusick 2001). Sequence-based approaches test functional variants in known genes across the genome and a map-based method systematically tests a dense map of anonymous polymorphisms across the genome. For the map-based study, it is critical to define the exact pattern of genome-wide LD. The primary limitation of this approach is the lack of complete disequilibrium with any tested candidate allele, which can substantially reduce power (Muller-Myhsok and Abel 1997). As genome-wide LD maps and gene- and population-specific data become available, it may be possible to evaluate potential haplotypes and to select most useful SNP markers. The human population is predicted to possess millions of SNPs, making it necessary to prioritise the choice of marker SNPs. Prior knowledge of the location of the SNPs, the LD pattern around the SNP and the allele frequencies of the SNPs allows for sophisticated study designs maximising the likelihood of successful outcome. The development of improved genotyping methods, possibly those based on a DNA pooling approach (Arnheim et al. 1985; Sasaki et al. 2001; Matyas et al. 2002; Werner et al. 2002), is required before genome-wide association studies are achievable at acceptable genotyping costs and effort. The investigation of the genetic basis of complex disorders has also set new challenges for statistical analysis and interpretation of the results. Statistical tools that can deal with large numbers of genotypes, complex disease inheritance and multiple phenotypes are required. Identification of effective gene mapping strategies is of primary importance for the future of complex disorders. Haplotype mapping focusing on different human populations (Daly et al. 2001; Gabriel et al. 2002) and systematic profiling of protein expression (Agaton et al. 2003) may facilitate the identification of novel genes and pathways involved in complex disease.

The identification of causative genes of complex disorders has proven to be challenging. Linkage and association studies have frequently produced conflicting results in ethnically distinct populations. However, recent advances in the genetics of complex diseases support an optimistic view of future prospects. Examples of recent success in the field of complex disease genetics include the identification of associations between NOD2 variants and Crohn's disease (Hugot et al. 2001; Ogura et al. 2001), and between

ADAM33 gene and asthma and bronchial hyperresponsiveness (Van Eerdewegh et al. 2002). Variation in the calpain-10 gene was initially linked and associated with type 2 diabetes mellitus susceptibility in a population of Mexican Americans (Horikawa et al. 2000). Subsequent association and linkage studies of the variants in this locus in other populations produced conflicting results (Evans et al. 2001; Fingerlin et al. 2002; Orho-Melander et al. 2002). However, a recent meta-analysis and a large association study confirm the role for calpain-10 variation in type 2 diabetes susceptibility (Weedon et al. 2003). It also demonstrates that meta-analyses of published genetic associations, combined with large replication studies, are a powerful approach to detect causative variants in common disease (Weedon et al. 2003).

Demonstrating the definitive proof of specific genetic causation in complex disease is elusive. A recent review suggests guidelines for steps to be taken to address the possible confounders of association before polymorphisms may be called 'causative' (Page et al. 2003). All known sources of bias should be eliminated before causative association is claimed. Once a haplotype associated with increased susceptibility to a disease has been identified, all SNPs that define the haplotype are candidates for causality. Gene expression analysis and complementary studies using animal models can aid the detection of the causal variant.

Genetic information may allow customising drug selection and dose based on each person's genetic profile. Such individualised treatment plans may avoid side effects, increase the specificity and efficiency of the treatments and potentially lead to pharmaco-economic gain. The identification of causative variants in genes for complex diseases allows the rational development of drugs to target the genes, protein pathways and networks demonstrated to be involved in primary disease pathogenesis (Collins et al. 2003). Unravelling of the biological function of the disease associated genetic variants and development of novel specific pharmacological treatments based on these findings, is a key challenge in the postgenomic era.

7.4 DIRECTIONS FOR FUTURE WORK IN THE GENETICS OF AS

A key objective for future efforts in the genetics of AS will be the transition from linkage analysis and modelling into identification of susceptibility genes and analysis of the pathways and mechanisms of AS. Large patient collections, the development of powerful analytical methods and high throughput SNP genotyping technologies will facilitate this endeavour. The recently established AS consortium (www.well.ox.ac.uk/brown/GASC.shtml) aims to promote collaborative research and makes a large patient collection available to establish future collaborative projects. Future meta-analyses of linkage projects undertaken by various groups is likely to have the high statistical power required to refine areas of linkage in AS.

Recent advances in the genetics of AS include the identification of an association in the susceptibility locus of the *IL-1* genetic cluster (McGarry et al. 2001b; van der Paardt et al. 2002a; Maksymowych et al. 2003). Genetic fine mapping studies are required to refine the definite causative variants involved in this locus. Further studies are necessary to establish the functional importance of these variants. Given the preliminary results that treatment with recombinant human IL-1Ra is effective in some patients with AS (Haibel et al. 2003), it may be possible that polymorphisms in this locus influence the outcome of this treatment. The involvement of cytokines in the pathogenesis of AS is likely to be multifaceted, and a variety of cytokine and cytokine receptor polymorphisms may interact with one another. Elucidation of the interaction of cytokine genes may provide insights into this poorly understood disease.

The development of novel research methods, such as the microarray technology, may shed light on the disease mechanisms involved in AS. The transcript pattern of affected and unaffected tissues can be compared by microarray technology, this could facilitate the identification of novel candidate genes in AS. The technology can identify genes that are dysregulated because of the susceptibility allele. Microarray analysis has the ability to assess the expression of a plethora of genes simultaneously. A significant weakness of microarrays is that they are prone to great experimental variability. However, the technology is relatively new, experimental procedures should improve and result in greater reliability in the future. Another disadvantage of microarrays in AS includes the difficulty of obtaining samples of the affected tissues, such as the spine, sacroiliac joints and enthesis. Peripheral blood, which is easy to obtain, may not be involved in the disease process, and peripheral blood cells may not exhibit disease-specific changes in gene expression. Recent microarray analysis demonstrated enhanced expression of monocyte chemotractant protein 1, proteasome subunit C2 and immunoglobulin heavy chain binding protein (BiP) in synovial fluid mononuclear cells of SpA patients, providing clues to the genes involved in synovial disease in SpA (Gu et al. 2002a). As BiP is a chaperone protein resident inside the ER, increases in BiP suggests that ER stress response due to accumulation of misfolded HLA-B27 may be involved in SpA (Gu et al. 2002b).

Two major types of microarrays have been developed: DNA and protein arrays. DNA microarrays allow the comparison of expression of mRNA between the affected and unaffected tissue. However, there is disparity between the relative expression levels of mRNA and their corresponding proteins. In addition, due to alternative splicing of mRNA and post-translational processes the number of functional proteins is known to exceed the number of its genes. Protein microarrays may increase the understanding of protein expression in AS and allow the profiling of relevant protein interactions. However, due to the diverse physical properties of proteins, the development of protein arrays is complicated and currently in its infancy. Data integration of DNA and protein microarrays, histology and genetics could facilitate the elucidation of disease pathways in AS and contribute to the stratification of individual patients into subgroups with an optimal response to specific therapeutic interventions.

The phenotype of a complex disease may encompass a wide range of clinical symptoms and subtypes. Thus, it is essential that standardised, validated ascertainment criteria are used in patient recruitment. The modified New York diagnostic criteria were used in this study (van der Linden et al. 1984a). However, the disease diagnosis arises from clinical traditions, and there are many phenotypic features in AS that cut across diagnostic categories such as sacroiliitis. AS patients vary in severity of symptoms and age of onset which results in heterogeneity in the phenotypic expression. This heterogeneity is likely to increase the number of loci involved and hamper the identification of disease genes. One approach is to subdivide the phenotype of AS into smaller and more homogeneous entities, which in turn may help in the identification of disease genes. This approach could involve for example the investigation of AS patient with associated IBD or AS patients with a younger age of symptom onset. An example of this kind of success was the initial finding of linkage for the breast cancer (*BRCA1*) gene. Genetic analysis of breast cancer families with early-onset disease yielded a significant linkage peak on chromosome 17q21, which allowed the identification of BRCA1 gene (Hall et al. 1990). Another commonly used approach is to use large multiply affected kindreds in which the affected individuals are likely to share more disease alleles (Tomfohrde et al. 1994). Such studies could facilitate genetic linkage studies in AS.

A number of linkage and association studies published to date are underpowered. Well-designed and large genetic studies are required to identify the non-*HLA-B27* genetic effects in AS. Collaborative projects may help to identify sufficiently large patient groups required for genetic analysis. The disadvantage of multi-centre studies is the potential for heterogeneity in the ascertainment and the phenotype definition of patients. Furthermore, the inclusion of samples from different ethnic populations may increase the heterogeneity of multi-centre studies as diverse populations may exhibit different LD patterns and disease associations. Ideally, a positive association and linkage signal is confirmed in several ethnically distinct populations and by several independent groups. Homogeneous populations such as the Finnish population may help us to define the genetic complexities of AS pathogenesis.

7.5 CONCLUDING REMARKS

This thesis is consistent with the concept that AS represents a disease with a complex genetic aetiology. A novel *HLA-DRB1* method was developed and emplyed in a large number of AS families to evaluate the contribution of *HLA-DRB1* alleles in AS.

Association between the MHC region and AS appears to be heterogenous, and genes within the MHC are involved both in disease susceptibility and the age of symptom onset. *HLA-B27* homozygosity increases susceptibility to AS. *TGFB1* gene polymorphisms do not appear to have a major impact in AS and they do not explain the observed linkage to chromosome 19.

The molecular mechanisms of disease pathogenesis of AS remain to be adequately defined. AS is a highly heritable disease and a large proportion of the genetic variance of the disease is not yet known. Unravelling these genetic variants could yield insights into the pathogenic mechanisms, allow the refinement of the molecular diagnosis of the disease and eventually open new avenues for tailored therapies.

REFERENCES

(1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. Cell 72:971-83

(1999) Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. Nature 401:921-3

Aaron S, Miller ML, Howard J, Fraser PA, Jackson JM, Larson MG, Glass DN (1985) Complementation with HLA-A and HLA-D locus alleles in ankylosing spondylitis with peripheral arthritis. J Rheumatol 12:553-7

Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66:279-92

Abecasis GR, Cookson WO (2000) GOLD--graphical overview of linkage disequilibrium. Bioinformatics 16:182-3

Abecasis GR, Noguchi E, Heinzmann A, Traherne JA, Bhattacharyya S, Leaves NI, Anderson GG, Zhang Y, Lench NJ, Carey A, Cardon LR, Moffatt MF, Cookson WO (2001) Extent and distribution of linkage disequilibrium in three genomic regions. Am J Hum Genet 68:191-197

Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97-101

Agaton C, Galli J, Hoiden Guthenberg I, Janzon L, Hansson M, Asplund A, Brundell E, Lindberg S, Ruthberg I, Wester K, Wurtz D, Hoog C, Lundeberg J, Stahl S, Ponten F, Uhlen M (2003) Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. Mol Cell Proteomics 2:405-14

Ahmad T, Neville M, Marshall SE, Armuzzi A, Mulcahy-Hawes K, Crawshaw J, Sato H, Ling KL, Barnardo M, Goldthorpe S, Walton R, Bunce M, Jewell DP, Welsh KI (2003) Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC. Hum Mol Genet 12:647-56

Aklillu E, Persson I, Bertilsson L, Johansson I, Rodrigues F, Ingelman-Sundberg M (1996) Frequent distribution of ultrarapid metabolizers of debrisoquine in an ethiopian population carrying duplicated and multiduplicated functional CYP2D6 alleles. J Pharmacol Exp Ther 278:441-6

Allcock RJ, Atrazhev AM, Beck S, De Jong PJ, Elliott JF, Forbes S, Halls K, Horton R, Osoegawa K, Rogers J, Sawcer S, Todd JA, Trowsdale J, Wang Y, Williams S (2002) The MHC haplotype project: A resource for HLA-linked association studies. Tissue Antigens 59:520-1

Allen RL, O'Callaghan CA, McMichael AJ, Bowness P (1999) Cutting edge: HLA-B27 can form a novel beta 2-microglobulin-free heavy chain homodimer structure. J Immunol 162:5045-8

Archer JR (1995) Ankylosing spondylitis, IgA, and transforming growth factors. Ann Rheum Dis 54:544-6

Ardlie K, Liu-Cordero SN, Eberle MA, Daly M, Barrett J, Winchester E, Lander ES, Kruglyak L (2001) Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. Am J Hum Genet 69:582-9

Ardlie KG, Kruglyak L, Seielstad M (2002a) Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3:299-309

Ardlie KG, Lunetta KL, Seielstad M (2002b) Testing for population subdivision and association in four case-control studies. Am J Hum Genet 71:304-11
Armas JB, Gonzalez S, Martinez-Borra J, Laranjeira F, Ribeiro E, Correia J, Ferreira ML, Toste M, Lopez-Vazquez A, Lopez-Larrea C (1999) Susceptibility to ankylosing spondylitis is independent of the Bw4 and Bw6 epitopes of HLA-B27 alleles. Tissue Antigens 53:237-43
Armstrong RD, Panayi GS, Welsh KI (1983) Histocompatibility antigens in psoriasis, psoriatic arthropathy, and ankylosing spondylitis. Ann Rheum Dis 42:142-6
Arnheim N, Strange C, Erlich H (1985) Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of the HLA class II loci. Proc Natl Acad Sci U S A 82:6970-4
Asumalahti K, Laitinen T, Itkonen-Vatjus R, Lokki ML, Suomela S, Snellman E, Saarialho-Kere U, Kere J (2000) A candidate gene for psoriasis near HLA-C, HCR (Pg8), is highly polymorphic with a disease-associated susceptibility allele. Hum Mol Genet 9:1533-42
Awad MR, El-Gamel A, Hasleton P, Turner DM, Sinnott PJ, Hutchinson IV (1998) Genotypic variation in the transforming growth factor-beta1 gene: association with transforming growth factor-beta1 production, fibrotic lung disease, and graft fibrosis after lung transplantation. Transplantation 66:1014-20
Badano JL, Katsanis N (2002) Beyond Mendel: an evolving view of human genetic disease transmission. Nat Rev Genet 3:779-89
Baer AN, McAllister CB, Wilkinson GR, Woosley RL, Pincus T (1986) Altered distribution of debrisoquine oxidation phenotypes in patients with systemic lupus erythematosus. Arthritis Rheum 29:843-50
Baeten D, Van Damme N, Van den Bosch F, Kruithof E, De Vos M, Mielants H, Veys EM, De Keyser F (2001) Impaired Th1 cytokine production in spondyloarthropathy is restored by anti-TNFalpha. Ann Rheum Dis 60:750-5
Baeten D, De Keyser F, Mielants H, Veys EM (2002) Immune linkages between inflammatory bowel disease and spondyloarthropathies. Curr Opin Rheumatol 14:342-7
Barnes PJ, Karin M (1997) Nuclear factor-kappaB: a pivotal transcription factor in chronic inflammatory diseases. N Engl J Med 336:1066-71
Benjamin R, Parham P (1990) Guilt by association: HLA-B27 and ankylosing spondylitis. Immunol Today 11:137-42
Beyeler C, Armstrong M, Bird HA, Idle JR, Daly AK (1996) Relationship between genotype for the cytochrome P450 CYP2D6 and susceptibility to ankylosing spondylitis and rheumatoid arthritis. Ann Rheum Dis 55:66-8
Blakemore AI, Watson PF, Weetman AP, Duff GW (1995) Association of Graves' disease with an allele of the interleukin-1 receptor antagonist gene. J Clin Endocrinol Metab 80:111-5
Blobe GC, Schiemann WP, Lodish HF (2000) Role of transforming growth factor beta in human disease. N Engl J Med 342:1350-8
Boehnke M (2000) A look at linkage disequilibrium. Nat Genet 25:246-7
Border WA, Noble NA (1994) Transforming growth factor beta in tissue fibrosis. N Engl J Med 331:1286-92
Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 33 Suppl:228-37

Bottazzo GF, Pujol-Borrell R, Hanafusa T, Feldmann M (1983) Role of aberrant HLA-DR expression and antigen presentation in induction of endocrine autoimmunity. Lancet 2:1115-9

Boyle LH, Hill Gaston JS (2003) Breaking the rules: the unconventional recognition of HLA-B27 by CD4+ T lymphocytes as an insight into the pathogenesis of the spondyloarthropathies. Rheumatology (Oxford) 42:404-12

Brandt J, Haibel H, Cornely D, Golder W, Gonzalez J, Reddig J, Thriene W, Sieper J, Braun J (2000) Successful treatment of active ankylosing spondylitis with the anti-tumor necrosis factor alpha monoclonal antibody infliximab. Arthritis Rheum 43:1346-52

Braun J, Bollow M, Eggens U, Konig H, Distler A, Sieper J (1994) Use of dynamic magnetic resonance imaging with fast imaging in the detection of early and advanced sacroiliitis in spondylarthropathy patients. Arthritis Rheum 37:1039-45

Braun J, Bollow M, Neure L, Seipelt E, Seyrekbasan F, Herbst H, Eggens U, Distler A, Sieper J (1995) Use of immunohistologic and in situ hybridization techniques in the examination of sacroiliac joint biopsy specimens from patients with ankylosing spondylitis. Arthritis Rheum 38:499-505

Braun J, Bollow M, Remlinger G, Eggens U, Rudwaleit M, Distler A, Sieper J (1998) Prevalence of spondylarthropathies in HLA-B27 positive and negative blood donors. Arthritis Rheum 41:58-67

Brewerton DA, Caffrey M, Nicholls A, D. W, James DC (1973a) Acute anterior uveitis and HLA-B27. Lancet 2:994-6

Brewerton DA, Hart FD, Nicholls A, Caffrey M, James DC, Sturrock RD (1973b) Ankylosing spondylitis and HL-A 27. Lancet 1:904-7

Broly F, Gaedigk A, Heim M, Eichelbaum M, Morike K, Meyer UA (1991) Debrisoquine/sparteine hydroxylation genotype and phenotype: analysis of common mutations and alleles of CYP2D6 in a European population. DNA Cell Biol 10:545-58

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63:861-9

Brooks BM, Flanagan BF, Thomas AL, Coleman JW (2001) Penicillin conjugates to interferon-gamma and reduces its activity: a novel drug-cytokine interaction. Biochem Biophys Res Commun 288:1175-81

Brophy S, Calin A (2001) Ankylosing spondylitis: interaction between genes, joints, age at onset, and disease expression. J Rheumatol 28:2283-8

Brown M, Bunce M, Calin A, Darke C, Wordsworth P (1996a) HLA-B associations of HLA-B27 negative ankylosing spondylitis: comment on the article by Yamaguchi et al. Arthritis Rheum 39:1768-9

Brown MA, Pile KD, Kennedy LG, Calin A, Darke C, Bell J, Wordsworth BP, Cornelis F (1996b) HLA class I associations of ankylosing spondylitis in the white population in the United Kingdom. Ann Rheum Dis 55:268-70

Brown MA, Jepson A, Young A, Whittle HC, Greenwood BM, Wordsworth BP (1997a) Ankylosing spondylitis in West Africans--evidence for a non-HLA-B27 protective effect. Ann Rheum Dis 56:68-70

Brown MA, Kennedy LG, MacGregor AJ, Darke C, Duncan E, Shatford JL, Taylor A, Calin A, Wordsworth P (1997b) Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and the environment. Arthritis Rheum 40:1823-8

Brown MA, Kennedy LG, Darke C, Gibson K, Pile KD, Shatford JL, Taylor A, Calin A, Wordsworth BP (1998a) The effect of HLA-DR genes on susceptibility to and severity of ankylosing spondylitis. Arthritis Rheum 41:460-5

Brown MA, Pile KD, Kennedy LG, Campbell D, Andrew L, March R, Shatford JL, Weeks DE, Calin A, Wordsworth BP (1998b) A genome-wide screen for susceptibility loci in ankylosing spondylitis. Arthritis Rheum 41:588-95

Brown MA, Edwards S, Hoyle E, Campbell S, Laval S, Daly AK, Pile KD, Calin A, Ebringer A, Weeks DE, Wordsworth BP (2000a) Polymorphisms of the CYP2D6 gene increase susceptibility to ankylosing spondylitis. Hum Mol Genet 9:1563-6

Brown MA, Laval SH, Brophy S, Calin A (2000b) Recurrence risk modelling of the genetic susceptibility to ankylosing spondylitis. Ann Rheum Dis 59:883-6

Brown MA, Brophy S, Bradbury L, Hamersma J, Timms A, Laval S, Cardon L, Calin A, Wordsworth BP (2003) Identification of major loci controlling clinical manifestations of ankylosing spondylitis. Arthritis Rheum 48:2234-9

Bruges Armas J, Laranjeira F, Peixoto M, Garrett F, Pimentel F (2000a) No increase in HLA class II alleles in B27+ patients. The Journal of Rheumatology 27:P1.3

Bruges Armas J, Peixoto M, Laranjeira F, Garrett F, Mendonca D (2000b) HLA class I and II alleles in B27 negative spondylarthritis (SpA) patients: is DR1 disease associated? The Journal of Rheumatology 27:P1.2

Bunce M, O'Neill CM, Barnardo MC, Krausa P, Browning MJ, Morris PJ, Welsh KI (1995) Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). Tissue Antigens 46:355-67

Burney RO, Pile KD, Gibson K, Calin A, Kennedy LG, Sinnott PJ, Powis SH, Wordsworth BP (1994) Analysis of the MHC class II encoded components of the HLA class I antigen processing pathway in ankylosing spondylitis. Ann Rheum Dis 53:58-60

Buxton SE, Benjamin RJ, Clayberger C, Parham P, Krensky AM (1992) Anchoring pockets in human histocompatibility complex leukocyte antigen (HLA) class I molecules: analysis of the conserved B pocket of HLA-B27. J Exp Med 175:809-20

Calin A, Marder A, Becks E, Burns T (1983) Genetic differences between B27 positive patients with ankylosing spondylitis and B27 positive healthy controls. Arthritis Rheum 26:1460-4

Calin A, Elswood J (1988) The relationship between pelvic, spinal and hip involvement in ankylosing spondylitis--one disease process or several? Br J Rheumatol 27:393-5

Calin A, Kennedy LG, Edmunds L, Will R (1993) Familial versus sporadic ankylosing spondylitis. Two different diseases? Arthritis Rheum 36:676-81

Calin A, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, Jenkinson T (1994) A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 21:2281-5

Calin A, Taurog JD (1998) The spondylarthritides. Oxford University Press

Cambien F, Ricard S, Troesch A, Mallet C, Generenaz L, Evans A, Arveiler D, Luc G, Ruidavets JB, Poirier O (1996) Polymorphisms of the transforming growth factor-beta 1 gene in relation to myocardial infarction and blood pressure. The Etude Cas-Temoin de l'Infarctus du Myocarde (ECTIM) Study. Hypertension 28:881-7

Cardon LR, Bell JI (2001) Association study designs for complex diseases. Nat Rev Genet 2:91-9
Carter N, Williamson L, Kennedy LG, Brown MA, Wordsworth BP (2000) Susceptibility to ankylosing spondylitis. Rheumatology (Oxford) 39:445
Cauli A, Dessole G, Fiorillo MT, Vacca A, Mameli A, Bitti P, Passiu G, Sorrentino R, Mathieu A (2002) Increased level of HLA-B27 expression in ankylosing spondylitis patients compared with healthy HLA-B27-positive subjects: a possible further susceptibility factor for the development of disease. Rheumatology (Oxford) 41:1375-9
Centrella M, McCarthy TL, Canalis E (1991) Transforming growth factor-beta and remodeling of bone. J Bone Joint Surg Am 73:1418-28
Cheifetz S, Weatherbee JA, Tsang ML, Anderson JK, Mole JE, Lucas R, Massague J (1987) The transforming growth factor-beta system, a complex pattern of cross-reactive ligands and receptors. Cell 48:409-15
Chen W, Wahl SM (1999) Manipulation of TGF-beta to control autoimmune and chronic inflammatory diseases. Microbes Infect 1:1367-80
Chen WM, Deng HW (2001) A general and accurate approach for computing the statistical power of the transmission disequilibrium test for complex disease genes. Genet Epidemiol 21:53-67
Christensen PM, Gotzsche PC, Brosen K (1998) The sparteine/debrisoquine (CYP2D6) oxidation polymorphism and the risk of Parkinson's disease: a meta-analysis. Pharmacogenetics 8:473-9
Clark AG (1990) Inference of haplotypes from PCR-amplified samples of diploid populations. Mol Biol Evol 7:111-22
Clayton D (1999) A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. Am J Hum Genet 65:1170-7
Colbert RA (2000) HLA-B27 misfolding: a solution to the spondyloarthropathy conundrum? Mol Med Today 6:224-30
Collado-Escobar MD, Nieto A, Mataran L, Raya E, Martin J (2000) Interleukin 6 gene promoter polymorphism is not associated with ankylosing spondylitis. J Rheumatol 27:1461-3
Collins A, Lonjou C, Morton NE (1999) Genetic epidemiology of single-nucleotide polymorphisms. Proc Natl Acad Sci U S A 96:15173-7
Collins FS, Morgan M, Patrinos A (2003) The Human Genome Project: lessons from large-scale biology. Science 300:286-90
Conneally PM (2003) The complexity of complex diseases. Am J Hum Genet 72:229-32
Cowling P, Ebringer R, Ebringer A (1980) Association of inflammation with raised serum IgA in ankylosing spondylitis. Ann Rheum Dis 39:545-9
Crane AM, Bradbury L, van Heel DA, McGovern DP, Brophy S, Rubin L, Siminovitch KA, Wordsworth BP, Calin A, Brown MA (2002) Role of NOD2 variants in spondylarthritis. Arthritis Rheum 46:1629-33
Crilly A, Hamilton J, Clark CJ, Jardine A, Madhok R (2002) Analysis of transforming growth factor beta1 gene polymorphisms in patients with systemic sclerosis. Ann Rheum Dis 61:678-81
Czarniecki CW, Chiu HH, Wong GH, McCabe SM, Palladino MA (1988) Transforming growth factor-beta 1 modulates the expression of class II histocompatibility antigens on human cells. J Immunol 140:4217-23
Dahlman I, Eaves IA, Kosoy R, Morrison VA, Heward J, Gough SC, Allahabadia A, Franklyn JA, Tuomilehto J, Tuomilehto-Wolf E, Cucca F, Guja C, Ionescu-

Tirgoviste C, Stevens H, Carr P, Nutland S, McKinney P, Shield JP, Wang W, Cordell HJ, Walker N, Todd JA, Concannon P (2002) Parameters for reliable results in genetic association studies in common disease. Nat Genet 30:149-50
Daly AK, Armstrong M, Monkman SC, Idle ME, Idle JR (1991) Genetic and metabolic criteria for the assignment of debrisoquine 4-hydroxylation (cytochrome P4502D6) phenotypes. Pharmacogenetics 1:33-41
Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. Nat Genet 29:229-32
D'Amato M (2002) The Crohn's associated NOD2 3020InsC frameshift mutation does not confer susceptibility to ankylosing spondylitis. J Rheumatol 29:2470-1
Danis VA, Millington M, Hyland VJ, Grennan D (1995) Cytokine production by normal human monocytes: inter-subject variation and relationship to an IL-1 receptor antagonist (IL-1Ra) gene polymorphism. Clin Exp Immunol 99:303-10
Daw EW, Thompson EA, Wijsman EM (2000) Bias in multipoint linkage analysis arising from map misspecification. Genet Epidemiol 19:366-80
de Juan MD, Reta A, Cancio J, Belzunegui J, Cuadrado E (1999) HLA-A*9, a probable secondary susceptibility marker to ankylosing spondylitis in Basque patients. Tissue Antigens 53:161-6
de la Chapelle A (1993) Disease gene mapping in isolated human populations: the example of Finland. J Med Genet 30:857-65
De La Concha EG, Fernandez-Arquero M, Martinez A, Vidal F, Vigil P, Conejero L, Garcia-Rodriguez MC, Fontan G (1999) HLA class II homozygosity confers susceptibility to common variable immunodeficiency (CVID). Clin Exp Immunol 116:516-20
de Vlam K, Mielants H, Cuvelier C, De Keyser F, Veys EM, De Vos M (2000) Spondyloarthropathy is underestimated in inflammatory bowel disease: prevalence and HLA association. J Rheumatol 27:2860-5
Derynck R, Choy L (1998) Transforming Growth Factor-beta and its receptors. In: Thomson AW (ed) The Cytokine Handbook. Academic Press Limited, pp 593-636
Devlin B, Roeder K, Wasserman L (2001) Genomic control, a new approach to genetic-based association studies. Theor Popul Biol 60:155-66
Djouadi K, Nedelec B, Tamouza R, Genin E, Ramasawmy R, Charron D, Delpech M, Laoussadi S (2001) Interleukin 1 gene cluster polymorphisms in multiplex families with spondylarthropathies. Cytokine 13:98-103
Dougados M, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, Cats A, Dijkmans B, Olivieri I, Pasero G, et al. (1991) The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. Arthritis Rheum 34:1218-27
Du X, Williams DA (1997) Interleukin-11: review of molecular, cell biology, and clinical use. Blood 89:3897-908
Dubois CM, Laprise MH, Blanchette F, Gentry LE, Leduc R (1995) Processing of transforming growth factor beta 1 precursor by human furin convertase. J Biol Chem 270:10618-24
Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Cucca F, Todd JA (2000) The genetically isolated populations of Finland and sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. Nat Genet 25:320-3
Ebringer A, Wilson C (2000) HLA molecules, bacteria and autoimmunity. J Med Microbiol 49:305-11

Edmunds L, Elswood J, Kennedy LG, Calin A (1991) Primary ankylosing spondylitis, psoriatic and enteropathic spondyloarthropathy: a controlled analysis. J Rheumatol 18:696-8

El Maghraoui A, Borderie D, Cherruau B, Edouard R, Dougados M, Roux C (1999) Osteoporosis, body composition, and bone turnover in ankylosing spondylitis. J Rheumatol 26:2205-9

Evans JC, Frayling TM, Cassell PG, Saker PJ, Hitman GA, Walker M, Levy JC, et al. (2001) Studies of association between the gene for calpain-10 and type 2 diabetes mellitus in the United Kingdom. Am J Hum Genet 69:544-52

Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921-7

Ezquerra A, Bragado R, Vega MA, Strominger JL, Woody J, Lopez de Castro JA (1985) Primary structure of papain-solubilized human histocompatibility antigen HLA-B27. Biochemistry 24:1733-41

Falk CT, Rubinstein P (1987) Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Ann Hum Genet 51 (Pt 3):227-33

Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. Jama 278:1349-56

Feldtkeller E, Khan MA, Van Der Heijde D, Van Der Linden S, Braun J (2003) Age at disease onset and diagnosis delay in HLA-B27 negative vs. positive patients with ankylosing spondylitis. Rheumatol Int 23:61-6

Ferreiros-Vidal I, Amarelo J, Barros F, Carracedo A, Gomez-Reino JJ, Gonzalez A (2003) Lack of association of ankylosing spondylitis with the most common NOD2 susceptibility alleles to Crohn's disease. J Rheumatol 30:102-4

Fingerlin TE, Erdos MR, Watanabe RM, Wiles KR, Stringham HM, Mohlke KL, Silander K, Valle TT, Buchanan TA, Tuomilehto J, Bergman RN, Boehnke M, Collins FS (2002) Variation in three single nucleotide polymorphisms in the calpain-10 gene not associated with type 2 diabetes in a large Finnish cohort. Diabetes 51:1644-8

Fiorillo MT, Cauli A, Carcassi C, Bitti PP, Vacca A, Passiu G, Bettosini F, Mathieu A, Sorrentino R (2003) Two distinctive HLA haplotypes harbor the B27 alleles negatively or positively associated with ankylosing spondylitis in Sardinia: implications for disease pathogenesis. Arthritis Rheum 48:1385-9

Flato B, Smerdel A, Johnston V, Lien G, Dale K, Vinje O, Egeland T, Sorskaar D, Forre O (2002) The influence of patient characteristics, disease variables, and HLA alleles on the development of radiographically evident sacroiliitis in juvenile idiopathic arthritis. Arthritis Rheum 46:986-94

Fraile A, Nieto A, Mataran L, Martin J (1998a) HSP70 gene polymorphisms in ankylosing spondylitis. Tissue Antigens 51:382-5

Fraile A, Nieto A, Vinasco J, Beraun Y, Martin J, Mataran L (1998b) Association of large molecular weight proteasome 7 gene polymorphism with ankylosing spondylitis. Arthritis Rheum 41:560-2

Fraile A, Collado MD, Mataran L, Martin J, Nieto A (2000) TAP1 and TAP2 polymorphism in Spanish patients with ankylosing spondylitis. Exp Clin Immunogenet 17:199-204

Francois RJ, Gardner DL, Degrave EJ, Bywaters EG (2000) Histopathologic evidence that sacroiliitis in ankylosing spondylitis is not merely enthesitis. Arthritis Rheum 43:2011-24

Francois RJ, Braun J, Khan MA (2001) Entheses and enthesitis: a histopathologic review and relevance to spondyloarthritides. Curr Opin Rheumatol 13:255-64

Fulker DW, Cherny SS, Sham PC, Hewitt JK (1999) Combined linkage and association sib-pair analysis for quantitative traits. Am J Hum Genet 64:259-67

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225-9

Garcia-Fernandez S, Gonzalez S, Mina Blanco A, Martinez-Borra J, Blanco-Gelaz M, Lopez-Vazquez A, Lopez-Larrea C (2001) New insights regarding HLA-B27 diversity in the Asian population. Tissue Antigens 58:259-62

Garcia-Gonzalez MA, Crusius JB, Strunk MH, Bouma G, Perez-Centeno CM, Pals G, Meuwissen SG, Pena AS (2000) TGFB1 gene polymorphisms and inflammatory bowel disease. Immunogenetics 51:869-72

Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A (1994) A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 21:2286-91

Gillard EF, Chamberlain JS, Murphy EG, Duff CL, Smith B, Burghes AH, Thompson MW, Sutherland J, Oss I, Bodrug SE, et al. (1989) Molecular and phenotypic analysis of patients with deletions within the deletion-rich region of the Duchenne muscular dystrophy (DMD) gene. Am J Hum Genet 45:507-20

Giltay R, Kostka G, Timpl R (1997) Sequence and expression of a novel member (LTBP-4) of the family of latent transforming growth factor-beta binding proteins. FEBS Lett 411:164-8

Goddard KA, Wijsman EM (2002) Characteristics of genetic markers and maps for cost-effective genome screens using diallelic markers. Genet Epidemiol 22:205-20

Goedecke V, Crane AM, Jaakkola E, Kaluza W, Laiho K, Weeks DE, Wilson J, Kauppi M, Kaarela K, Tuomilehto J, Wordsworth BP, Brown MA (2003) Interleukin 10 polymorphisms in ankylosing spondylitis. Genes Immun 4:74-6

Gonzalez FJ, Nebert DW (1990) Evolution of the P450 gene superfamily: animal-plant 'warfare', molecular drive and human genetic differences in drug oxidation. Trends Genet 6:182-6

Gonzalez-Roces S, Alvarez MV, Gonzalez S, Dieye A, Makni H, Woodfield DG, Housan L, Konenkov V, Abbadi MC, Grunnet N, Coto E, Lopez-Larrea C (1997) HLA-B27 polymorphism and worldwide susceptibility to ankylosing spondylitis. Tissue Antigens 49:116-23

Goring HH, Terwilliger JD, Blangero J (2001) Large upward bias in estimation of locus-specific effects from genomewide scans. Am J Hum Genet 69:1357-69

Gough AC, Miles JS, Spurr NK, Moss JE, Gaedigk A, Eichelbaum M, Wolf CR (1990) Identification of the primary gene defect at the cytochrome P450 CYP2D locus. Nature 347:773-6

Grainger DJ, Heathcote K, Chiano M, Snieder H, Kemp PR, Metcalfe JC, Carter ND, Spector TD (1999) Genetic control of the circulating concentration of transforming growth factor type beta1. Hum Mol Genet 8:93-7

Gran JT, Husby G, Hordvik M (1985) Prevalence of ankylosing spondylitis in males and females in a young middle-aged population of Tromso, northern Norway. Ann Rheum Dis 44:359-67

Gran JT, Ostensen M (1998) Spondyloarthritides in females. Baillieres Clin Rheumatol 12:695-715

Gratacos J, Collado A, Filella X, Sanmarti R, Canete J, Llena J, Molina R, Ballesta A, Munoz-Gomez J (1994) Serum cytokines (IL-6, TNF-alpha, IL-1 beta and IFN-gamma) in ankylosing spondylitis: a close correlation between serum IL-6 and disease activity and severity. Br J Rheumatol 33:927-31

Green PM, Saad S, Lewis CM, Giannelli F (1999) Mutation rates in humans. I. Overall and sex-specific rates obtained from a population study of hemophilia B. Am J Hum Genet 65:1572-9

Griem P, Wulferink M, Sachs B, Gonzalez JB, Gleichmann E (1998) Allergic and autoimmune reactions to xenobiotics: how do they arise? Immunol Today 19:133-41

Groh V, Bahram S, Bauer S, Herman A, Beauchamp M, Spies T (1996) Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. Proc Natl Acad Sci U S A 93:12445-50

Groh V, Steinle A, Bauer S, Spies T (1998) Recognition of stress-induced MHC molecules by intestinal epithelial gammadelta T cells. Science 279:1737-40

Gu J, Marker-Hermann E, Baeten D, Tsai WC, Gladman D, Xiong M, Deister H, Kuipers JG, Huang F, Song YW, Maksymowych W, Kalsi J, Bannai M, Seta N, Rihl M, Crofford LJ, Veys E, De Keyser F, Yu DT (2002a) A 588-gene microarray analysis of the peripheral blood mononuclear cells of spondyloarthropathy patients. Rheumatology (Oxford) 41:759-66

Gu J, Rihl M, Marker-Hermann E, Baeten D, Kuipers JG, Song YW, Maksymowych WP, Burgos-Vargas R, Veys EM, De Keyser F, Deister H, Xiong M, Huang F, Tsai WC, Yu DT (2002b) Clues to pathogenesis of spondyloarthropathy derived from synovial fluid mononuclear cell gene expression profiles. J Rheumatol 29:2159-64

Guo HC, Madden DR, Silver ML, Jardetzky TS, Gorga JC, Strominger JL, Wiley DC (1993) Comparison of the P2 specificity pocket in three human histocompatibility antigens: HLA-A*6801, HLA-A*0201, and HLA-B*2705. Proc Natl Acad Sci U S A 90:8053-7

Hackstein H, Morelli AE, Thomson AW (2001) Designer dendritic cells for tolerance induction: guided not misguided missiles. Trends Immunol 22:437-42

Haibel H, Rudwaleit M, Braun J, Listing J, Sieper J (2003) Preliminary results of an open study with anakinra in active ankylosing spondylitis over 24 weeks. Ann Rheum Dis 62:249

Haldane JBS (1919) The combination of linkage values, and the calculation of distances between the loci of linked factors. J Genet 8:299-309

Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250:1684-9

Hamersma J, Cardon LR, Bradbury L, Brophy S, van der Horst-Bruinsma I, Calin A, Brown MA (2001) Is disease severity in ankylosing spondylitis genetically determined? Arthritis Rheum 44:1396-400

Hammer RE, Maika SD, Richardson JA, Tang JP, Taurog JD (1990) Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human beta

2m: an animal model of HLA-B27-associated human disorders. Cell 63:1099-112

Hartl D, Clark A (1997) Principles of population genetics. Sinauer Associates

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3-19

Hastbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat Genet 2:204-11

Hastbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, Hamilton BA, Kusumi K, Trivedi B, Weaver A, et al. (1994) The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. Cell 78:1073-87

Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. Genetics 117:331-41

Heresbach D, Alizadeh M, Dabadie A, Le Berre N, Colombel JF, Yaouanq J, Bretagne JF, Semana G (1997) Significance of interleukin-1beta and interleukin-1 receptor antagonist genetic polymorphism in inflammatory bowel diseases. Am J Gastroenterol 92:1164-9

Hermann E, Yu DT, Meyer zum Buschenfelde KH, Fleischer B (1993) HLA-B27-restricted CD8 T cells derived from synovial fluids of patients with reactive arthritis and ankylosing spondylitis. Lancet 342:646-50

Hill WG, Robertson AR (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226-231

Ho AM, Johnson MD, Kingsley DM (2000) Role of the mouse ank gene in control of tissue calcification and arthritis. Science 289:265-70

Hoffman IE, Demetter P, Peeters M, De Vos M, Mielants H, Veys EM, De Keyser F (2003) Anti-Saccharomyces cerevisiae IgA antibodies are raised in ankylosing spondylitis and undifferentiated spondyloarthropathy. Ann Rheum Dis 62:455-9

Hohler T, Schaper T, Schneider PM, Meyer zum Buschenfelde KH, Marker-Hermann E (1998) Association of different tumor necrosis factor alpha promoter allele frequencies with ankylosing spondylitis in HLA-B27 positive individuals. Arthritis Rheum 41:1489-92

Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, del Bosque-Plata L, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet 26:163-75

Hoyle E, Laval SH, Calin A, Wordsworth BP, Brown MA (2000) The X-chromosome and susceptibility to ankylosing spondylitis. Arthritis Rheum 43:1353-5

Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411:599-603

Hurme M, Santtila S (1998) IL-1 receptor antagonist (IL-1Ra) plasma levels are co-ordinately regulated by both IL-1Ra and IL-1beta genes. Eur J Immunol 28:2598-602

Hurme M, Lahdenpohja N, Santtila S (1998) Gene polymorphisms of interleukins 1 and 10 in infectious and autoimmune diseases. Ann Med 30:469-73

Ingelman-Sundberg M, Oscarson M, McLellan RA (1999) Polymorphic human cytochrome P450 enzymes: an opportunity for individualized drug treatment. Trends Pharmacol Sci 20:342-9

Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG (2001) Replication validity of genetic association studies. Nat Genet 29:306-9

Janssens K, Gershoni-Baruch R, Guanabens N, Migone N, Ralston S, Bonduelle M, Lissens W, Van Maldergem L, Vanhoenacker F, Verbruggen L, Van Hul W (2000) Mutations in the gene encoding the latency-associated peptide of TGF-beta 1 cause Camurati-Engelmann disease. Nat Genet 26:273-5

Jawaheer D, Li W, Graham RR, Chen W, Damle A, Xiao X, Monteiro J, Khalili H, Lee A, Lundsten R, Begovich A, Bugawan T, Erlich H, Elder JT, Criswell LA, Seldin MF, Amos CI, Behrens TW, Gregersen PK (2002) Dissecting the genetic complexity of the association between human leukocyte antigens and rheumatoid arthritis. Am J Hum Genet 71:585-94

Johansson I, Lundqvist E, Bertilsson L, Dahl ML, Sjoqvist F, Ingelman-Sundberg M (1993) Inherited amplification of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine. Proc Natl Acad Sci U S A 90:11825-9

Johnson DW, Berg JN, Baldwin MA, Gallione CJ, Marondel I, Yoon SJ, Stenzel TT, Speer M, Pericak-Vance MA, Diamond A, Guttmacher AE, Jackson CE, Attisano L, Kucherlapati R, Porteous ME, Marchuk DA (1996) Mutations in the activin receptor-like kinase 1 gene in hereditary haemorrhagic telangiectasia type 2. Nat Genet 13:189-95

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA (2001) Haplotype tagging for the identification of common disease genes. Nat Genet 29:233-7

Jones SD, Steiner A, Garrett SL, Calin A (1996) The Bath Ankylosing Spondylitis Patient Global Score (BAS-G). Br J Rheumatol 35:66-71

Judson R, Salisbury B, Schneider J, Windemuth A, Stephens JC (2002) How many SNPs does a genome-wide haplotype map require? Pharmacogenomics 3:379-91

Kaartinen V, Warburton D (2003) Fibrillin controls TGF-beta activation. Nat Genet 33:331-2

Kagimoto M, Heim M, Kagimoto K, Zeugin T, Meyer UA (1990) Multiple mutations of the human cytochrome P450IID6 gene (CYP2D6) in poor metabolizers of debrisoquine. Study of the functional significance of individual mutations by expression of chimeric genes. J Biol Chem 265:17209-14

Kaijzel EL, Brinkman BM, van Krugten MV, Smith L, Huizinga TW, Verjans GM, Breedveld FC, Verweij CL (1999) Polymorphism within the tumor necrosis factor alpha (TNF) promoter region in patients with ankylosing spondylitis. Hum Immunol 60:140-4

Kaipiainen-Seppanen O, Aho K, Heliovaara M (1997) Incidence and prevalence of ankylosing spondylitis in Finland. J Rheumatol 24:496-9

Kaluza W, Leirisalo-Repo M, Marker-Hermann E, Westman P, Reuss E, Hug R, Mastrovic K, Stradmann-Bellinghausen B, Granfors K, Galle PR, Hohler T (2001) IL10.G microsatellites mark promoter haplotypes associated with protection against the development of reactive arthritis in Finnish patients. Arthritis Rheum 44:1209-14

Kamiya M, Harada A, Mizuno M, Iwata H, Yamada Y (2001) Association between a polymorphism of the transforming growth factor-beta1 gene and genetic susceptibility to ossification of the posterior longitudinal ligament in Japanese patients. Spine 26:1264-6; discussion 1266-7

Kennedy LG, Will R, Calin A (1993) Sex ratio in the spondyloarthropathies and its relationship to phenotypic expression, mode of inheritance and age at onset. J Rheumatol 20:1900-4

Kere J (2001) Human population genetics: lessons from Finland. Annu Rev Genomics Hum Genet 2:103-28

Khan MA, Kushner I, Braun WE (1977) Comparison of clinical features in HLA-B27 positive and negative patients with ankylosing spondylitis. Arthritis Rheum 20:909-12

Khan MA, Kushner I, Braun WE, Zachary AA, Steinberg AG (1978) HLA--B27 homozygosity in ankylosing spondylitis: relationship to risk and severity. Tissue Antigens 11:434-8

Khan MA, Khan MK (1982) Diagnostic value of HLA-B27 testing ankylosing spondylitis and Reiter's syndrome. Ann Intern Med 96:70-6

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073-80

Khare SD, Luthra HS, David CS (1995) Spontaneous inflammatory arthritis in HLA-B27 transgenic mice lacking beta 2-microglobulin: a model of human spondyloarthropathies. J Exp Med 182:1153-8

Kimber I, Dearman RJ (2002) Immunologic basis for autoimmunity and the potential influences of xenobiotics. Toxicol Lett 127:77-81

Kimura S, Umeno M, Skoda RC, Meyer UA, Gonzalez FJ (1989) The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene. Am J Hum Genet 45:889-904

Kinnear PR, Gray CD (2000) SPSS for windows made simple. Psychology Press, Sussex

Kinoshita A, Saito T, Tomita H, Makita Y, Yoshida K, Ghadami M, Yamada K, Kondo S, Ikegawa S, Nishimura G, Fukushima Y, Nakagomi T, Saito H, Sugimoto T, Kamegaya M, Hisa K, Murray JC, Taniguchi N, Niikawa N, Yoshiura K (2000) Domain-specific mutations in TGFB1 result in Camurati-Engelmann disease. Nat Genet 26:19-20

Kirkwood BR (1988) Essentials of medical statistics. Blackwell, Oxford

Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, Linnoila M, Goldman D, Long JC (1998) Dual origins of Finns revealed by Y chromosome haplotype variation. Am J Hum Genet 62:1171-9

Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM (1987) Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell 50:509-17

Kol A, Bourcier T, Lichtman AH, Libby P (1999) Chlamydial and human heat shock protein 60s activate human vascular endothelium, smooth muscle cells, and macrophages. J Clin Invest 103:571-7

Kollnberger S, Bird L, Sun MY, Retiere C, Braud VM, McMichael A, Bowness P (2002) Cell-surface expression and immune receptor recognition of HLA-B27 homodimers. Arthritis Rheum 46:2972-82

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179-88

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002) A high-resolution recombination map of the human genome. Nat Genet 31:241-7

Kortunay S, Bozkurt A, Bathum L, Basci NE, Calguneri M, Brosen K, Kayaalp SO (1999) CYP2D6 polymorphism in systemic lupus erythematosus patients. Eur J Clin Pharmacol 55:21-5

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172-5

Kotsch K, Wehling J, Blasczyk R (1999) Sequencing of HLA class II genes based on the conserved diversity of the non-coding regions: sequencing based typing of HLA-DRB genes. Tissue Antigens 53:486-97

Krimpenfort P, Rudenko G, Hochstenbach F, Guessow D, Berns A, Ploegh H (1987) Crosses of two independently derived transgenic mice demonstrate functional complementation of the genes encoding heavy (HLA-B27) and light (beta 2-microglobulin) chains of HLA class I antigens. Embo J 6:1673-6

Kruglyak L (1997) The use of a genetic map of biallelic markers in linkage studies. Nat Genet 17:21-4

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22:139-44

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347-63

La Nasa G, Mathieu A, Mulargia M, Carcassi C, Vacca A, Ledda A, Contu L (1993) Association of the HLA-A2, CW2, B27, S31, DR2 haplotype with ankylosing spondylitis. A possible role of non-B27 factors in the disease. Dis Markers 11:191-203

Laan M, Paabo S (1997) Demographic history and linkage disequilibrium in human populations. Nat Genet 17:435-8

Lahesmaa R, Skurnik M, Granfors K, Mottonen T, Saario R, Toivanen A, Toivanen P (1992) Molecular mimicry in the pathogenesis of spondyloarthropathies. A critical appraisal of cross-reactivity between microbial antigens and HLA-B27. Br J Rheumatol 31:221-9

Laitinen T, Daly MJ, Rioux JD, Kauppi P, Laprise C, Petays T, Green T, Cargill M, Haahtela T, Lander ES, Laitinen LA, Hudson TJ, Kere J (2001) A susceptibility locus for asthma-related traits on chromosome 7 revealed by genome-wide scan in a founder population. Nat Genet 28:87-91

Lalouel JM, White RL (1997) Analysis of genetic linkage. In: Rimoin DL, Connor JM, Pyeritz RE (eds) Emery and Rimoin's principles and practice of medical genetics. Churchill Livingstone, pp 111-125

Laloux L, Voisin MC, Allain J, Martin N, Kerboull L, Chevalier X, Claudepierre P (2001) Immunohistological study of entheses in spondyloarthropathies: comparison in rheumatoid arthritis and osteoarthritis. Ann Rheum Dis 60:316-21

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241-7

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409:860-921

Langdahl BL, Knudsen JY, Jensen HK, Gregersen N, Eriksen EF (1997) A sequence variation: 713-8delC in the transforming growth factor-beta 1 gene has higher prevalence in osteoporotic women than in normal women and is associated with very low bone mass in osteoporotic women and increased bone turnover in both osteoporotic and normal women. Bone 20:289-94

Lange U, Jung O, Teichmann J, Neeck G (2001) Relationship between disease activity and serum levels of vitamin D metabolites and parathyroid hormone in ankylosing spondylitis. Osteoporos Int 12:1031-5

Lathrop GM, Lalouel JM (1984) Easy calculations of lod scores and genetic risks on small computers. Am J Hum Genet 36:460-5

Lautermann D, Braun J (2002) Ankylosing spondylitis--cardiac manifestations. Clin Exp Rheumatol 20:S11-5

Laval SH, Bradbury L, Darke C, Brophy S, Calin A, Brown MA (2000) The role of HLA-DR genes in ankylosing spondylitis complicating inflammatory bowel disease. Rheumatology 39:64

Laval SH, Timms A, Edwards S, Bradbury L, Brophy S, Milicic A, Rubin L, Siminovitch KA, Weeks DE, Calin A, Wordsworth BP, Brown MA (2001) Whole-genome screening in ankylosing spondylitis: evidence of non-MHC genetic-susceptibility loci. Am J Hum Genet 68:918-26

Lee LG, Connell CR, Woo SL, Cheng RD, McArdle BF, Fuller CW, Halloran ND, Wilson RK (1992) DNA sequencing with dye-labeled terminators and T7 DNA polymerase: effect of dyes and dNTPs on incorporation of dye-terminators and probability analysis of termination fragments. Nucleic Acids Res 20:2471-83

Leirisalo-Repo M, Turunen U, Stenman S, Helenius P, Seppala K (1994) High frequency of silent inflammatory bowel disease in spondylarthropathy. Arthritis Rheum 37:23-31

Lemos MC, Cabrita FJ, Silva HA, Vivan M, Placido F, Regateiro FJ (1999) Genetic polymorphism of CYP2D6, GSTM1 and NAT2 and susceptibility to haematological neoplasias. Carcinogenesis 20:1225-9

Letterio JJ, Roberts AB (1998) Regulation of immune responses by TGF-beta. Annu Rev Immunol 16:137-61

Lewontin RC (1988) On measures of gametic disequilibrium. Genetics 120:849-52

Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33:177-82

Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W, Rubin EM, Frazer KA (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science 288:136-40

Lopez-Larrea C, Sujirachato K, Mehra NK, Chiewsilp P, Isarangkura D, Kanga U, Dominguez O, Coto E, Pena M, Setien F, et al. (1995) HLA-B27 subtypes in Asian patients with ankylosing spondylitis. Evidence for new associations. Tissue Antigens 45:169-76

Lopez-Larrea C, Gonzalez S, Martinez-Borra J (1998) The role of HLA-B27 polymorphism and molecular mimicry in spondylarthropathy. Mol Med Today 4:540-9

Lopez-Larrea C, Mijiyawa M, Gonzalez S, Fernandez-Morera JL, Blanco-Gelaz MA, Martinez-Borra J, Lopez-Vazquez A (2002) Association of ankylosing

spondylitis with HLA-B*1403 in a West African population. Arthritis Rheum 46:2968-71
Lympany PA, Avila JJ, Mullighan C, Marshall S, Welsh KI, du Bois RM (1998) Rapid genotyping of transforming growth factor beta1 gene polymorphisms in a UK Caucasoid control population using the polymerase chain reaction and sequence-specific primers. Tissue Antigens 52:573-8
Madden DR, Gorga JC, Strominger JL, Wiley DC (1991) The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. Nature 353:321-5
Maksymowych WP (2000) Ankylosing spondylitis--at the interface of bone and cartilage. J Rheumatol 27:2295-301
Maksymowych WP, Adlam N, Lind D, Russell AS (1997a) Polymorphism of the LMP2 gene and disease phenotype in ankylosing spondylitis: no association with disease severity. Clin Rheumatol 16:461-5
Maksymowych WP, Gorodezky C, Olivo A, Alaez C, Wong C, Burgos-Vargas R, Sanchez-Corona J, Ramos-Remus C, Russell AS (1997b) HLA-DRB1*08 influences the development of disease in Mexican Mestizo with spondyloarthropathy. J Rheumatol 24:904-7
Maksymowych WP, Jhangri GS, Gorodezky C, Luong M, Wong C, Burgos-Vargas R, Morenot M, Sanchez-Corona J, Ramos-Remus C, Russell AS (1997c) The LMP2 polymorphism is associated with susceptibility to acute anterior uveitis in HLA-B27 positive juvenile and adult Mexican subjects with ankylosing spondylitis. Ann Rheum Dis 56:488-92
Maksymowych WP, Tao S, Vaile J, Suarez-Almazor M, Ramos-Remus C, Russell AS (2000) LMP2 polymorphism is associated with extraspinal disease in HLA-B27 negative Caucasian and Mexican Mestizo patients with ankylosing spondylitis. J Rheumatol 27:183-9
Maksymowych WP, Reeve JP, Reveille JD, Akey JM, Buenviaje H, O'Brien L, Peloso PM, Thomson GT, Jin L, Russell AS (2003) High-throughput single-nucleotide polymorphism analysis of the IL1RN locus in patients with ankylosing spondylitis by matrix-assisted laser desorption ionization-time-of-flight mass spectrometry. Arthritis Rheum 48:2011-8
Martinez-Borra J, Gonzalez S, Lopez-Vazquez A, Gelaz MA, Armas JB, Kanga U, Mehra NK, Lopez-Larrea C (2000) HLA-B27 alone rather than B27-related class I haplotypes contributes to ankylosing spondylitis susceptibility. Hum Immunol 61:131-9
Marzo-Ortega H, McGonagle D, O'Connor P, Emery P (2001) Efficacy of etanercept in the treatment of the entheseal pathology in resistant spondylarthropathy: a clinical and magnetic resonance imaging study. Arthritis Rheum 44:2112-7
Massague J (1998) TGF-beta signal transduction. Annu Rev Biochem 67:753-91
Massague J (2000) How cells read TGF-beta signals. Nat Rev Mol Cell Biol 1:169-78
Matise TC, Sachidanandam R, Clark AG, Kruglyak L, Wijsman E, Kakol J, Buyske S, et al. (2003) A 3.9-centimorgan-resolution human single-nucleotide polymorphism linkage map and screening set. Am J Hum Genet 73:271-84
Matyas G, Giunta C, Steinmann B, Hossle JP, Hellwig R (2002) Quantification of single nucleotide polymorphisms: a novel method that combines primer extension assay and capillary electrophoresis. Hum Mutat 19:58-68
Mau W, Zeidler H, Mau R, Majewski A, Freyschmidt J, Stangel W, Deicher H (1988) Clinical features and prognosis of patients with possible ankylosing spondylitis. Results of a 10-year followup. J Rheumatol 15:1109-14

McAllister KA, Grogg KM, Johnson DW, Gallione CJ, Baldwin MA, Jackson CE, Helmbold EA, Markel DS, McKinnon WC, Murrell J, et al. (1994) Endoglin, a TGF-beta binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1. Nat Genet 8:345-51
McGarry F, Walker R, Sturrock R, Field M (1999) The -308.1 polymorphism in the promoter region of the tumor necrosis factor gene is associated with ankylosing spondylitis independent of HLA-B27. J Rheumatol 26:1110-6
McGarry F, Cousins L, Sturrock RD, Field M (2001a) A polymorphism within the transforming growth factor ß1 (TGFß1) gene is associated with ankylosing spondylitis (AS). Br J Rheumatol:16
McGarry F, Neilly J, Anderson N, Sturrock R, Field M (2001b) A polymorphism within the interleukin 1 receptor antagonist (IL-1Ra) gene is associated with ankylosing spondylitis. Rheumatology (Oxford) 40:1359-64
McGonagle D, Gibbon W, Emery P (1998a) Classification of inflammatory arthritis by enthesitis. Lancet 352:1137-40
McGonagle D, Gibbon W, O'Connor P, Green M, Pease C, Emery P (1998b) Characteristic magnetic resonance imaging entheseal changes of knee synovitis in spondylarthropathy. Arthritis Rheum 41:694-700
McGonagle D, Emery P (2000) Enthesitis, osteitis, microbes, biomechanics, and immune reactivity in ankylosing spondylitis. J Rheumatol 27:2302-4
McGonagle D, Stockwin L, Isaacs J, Emery P (2001) An enthesitis based model for the pathogenesis of spondyloarthropathy. additive effects of microbial adjuvant and biomechanical factors at disease sites. J Rheumatol 28:2155-9
McGonagle D, Marzo-Ortega H, Benjamin M, Emery P (2003) Report on the Second international Enthesitis Workshop. Arthritis Rheum 48:896-905
Mear JP, Schreiber KL, Munz C, Zhu X, Stevanovic S, Rammensee HG, Rowland-Jones SL, Colbert RA (1999) Misfolding of HLA-B27 as a result of its B pocket suggests a novel mechanism for its role in susceptibility to spondyloarthropathies. J Immunol 163:6665-70
Miceli-Richard C, Zouali H, Lesage S, Thomas G, Hugot JP, Said-Nahal R, Breban M (2002) CARD15/NOD2 analyses in spondylarthropathy. Arthritis Rheum 46:1405-6
Miehle W, Schattenkirchner M, Albert D, Bunge M (1985) HLA-DR4 in ankylosing spondylitis with different patterns of joint involvement. Ann Rheum Dis 44:39-44
Mielants H, Veys EM, Cuvelier C, De Vos M, Goemaere S, De Clercq L, Schatteman L, Gyselbrecht L, Elewaut D (1995) The evolution of spondyloarthropathies in relation to gut histology. III. Relation between gut and joint. J Rheumatol 22:2279-84
Mijiyawa M, Oniankitan O, Khan MA (2000) Spondyloarthropathies in sub-Saharan Africa. Curr Opin Rheumatol 12:281-6
Milicic A, Lindheimer F, Laval S, Rudwaleit M, Ackerman H, Wordsworth P, Hohler T, Brown MA (2000) Interethnic studies of TNF polymorphisms confirm the likely presence of a second MHC susceptibility locus in ankylosing spondylitis. Genes Immun 1:418-22
Modrek B, Lee C (2002) A genomic view of alternative splicing. Nat Genet 30:13-9
Monowarul Islam SM, Numaga J, Fujino Y, Masuda K, Ohda H, Hirata R, Maeda H, Mitsui H (1995) HLA-DR8 and acute anterior uveitis in ankylosing spondylitis. Arthritis Rheum 38:547-50

Mori K, Chano T, Ikeda T, Ikegawa S, Matsusue Y, Okabe H, Saeki Y (2003) Decrease in serum nucleotide pyrophosphatase activity in ankylosing spondylitis. Rheumatology (Oxford) 42:62-5

Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277-318

Mukhopadhyay N, Almasy L, Schroeder M, Mulvihill W, Weeks DE (1999) Mega2, a data-handling program for facilitating genetic linkage and association analyses. Am J Hum Genet 65:A436

Muller-Myhsok B, Abel L (1997) Genetic analysis of complex diseases. Science 275:1328-9; author reply 1329-30

Muro M, Moya-Quiles MR, Marin L, Torio A, Vallejo C, Moraleda JM, Alvarez-Lopez MR (2002) Report of recombinations between HLA loci within two families: utility of high resolution typing. Clin Transplant 16:329-33

Murray PI, Clay CD, Mappin C, Salmon M (1999) Molecular analysis of resolving immune responses in uveitis. Clin Exp Immunol 117:455-61

Nasution AR, Mardjuadi A, Kunmartini S, Suryadhana NG, Setyohadi B, Sudarsono D, Lardy NM, Feltkamp TE (1997) HLA-B27 subtypes positively and negatively associated with spondyloarthropathy. J Rheumatol 24:1111-4

Nebert DW (1997) Polymorphisms in drug-metabolizing enzymes: what is their clinical relevance and why do they exist? Am J Hum Genet 60:265-71

Neptune ER, Frischmeyer PA, Arking DE, Myers L, Bunton TE, Gayraud B, Ramirez F, Sakai LY, Dietz HC (2003) Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. Nat Genet 33:407-11

Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res 17:2503-16

Newton J, Brown MA, Milicic A, Ackerman H, Darke C, Wilson JN, Wordsworth BP, Kwiatkowski D (2003) The effect of HLA-DR on susceptibility to rheumatoid arthritis is influenced by the associated lymphotoxin alpha-tumor necrosis factor haplotype. Arthritis Rheum 48:90-6

Noda M, Camilliere JJ (1989) In vivo stimulation of bone formation by transforming growth factor-beta. Endocrinology 124:2991-4

Norton N, Williams NM, Williams HJ, Spurlock G, Kirov G, Morris DW, Hoogendoorn B, Owen MJ, O'Donovan MC (2002) Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. Hum Genet 110:471-8

Nurnberg P, Thiele H, Chandler D, Hohne W, Cunningham ML, Ritter H, Leschik G, Uhlmann K, Mischung C, Harrop K, Goldblatt J, Borochowitz ZU, Kotzot D, Westermann F, Mundlos S, Braun HS, Laing N, Tinschert S (2001) Heterozygous mutations in ANKH, the human ortholog of the mouse progressive ankylosis gene, result in craniometaphyseal dysplasia. Nat Genet 28:37-41

Nyholt DR (2000) All LODs are not created equal. Am J Hum Genet 67:282-8

Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nunez G, Cho JH (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 411:603-6

Oklu R, Hesketh R (2000) The latent transforming growth factor beta binding protein (LTBP) family. Biochem J 352 Pt 3:601-10

Olivieri I, Ciancio G, Padula A, Gaudiano C, Masciandaro S, Moro L, Durante E, Pozzi S, Ferrara GB (2002) The HLA-B*2709 subtype confers susceptibility to spondylarthropathy. Arthritis Rheum 46:553-4

Oostveen J, Prevo R, den Boer J, van de Laar M (1999) Early detection of sacroiliitis on magnetic resonance imaging and subsequent development of sacroiliitis on plain radiography. A prospective, longitudinal study. J Rheumatol 26:1953-8

Orho-Melander M, Klannemark M, Svensson MK, Ridderstrale M, Lindgren CM, Groop L (2002) Variants in the calpain-10 gene predispose to insulin resistance and elevated free fatty acid levels. Diabetes 51:2658-64

Page GP, George V, Go RC, Page PZ, Allison DB (2003) Are We There Yet?: deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. Am J Hum Genet 73:711-9

Pastinen T, Partanen J, Syvanen AC (1996) Multiplex, fluorescent, solid-phase minisequencing for efficient screening of DNA sequence variation. Clin Chem 42:1391-7

Payami H, Joe S, Farid NR, Stenszky V, Chan SH, Yeo PP, Cheah JS, Thomson G (1989) Relative predispositional effects (RPEs) of marker alleles with disease: HLA-DR alleles and Graves disease. Am J Hum Genet 45:541-6

Peltonen L, McKusick VA (2001) Genomics and medicine. Dissecting human disease in the postgenomic era. Science 291:1224-9

Pendleton A, Johnson MD, Hughes A, Gurley KA, Ho AM, Doherty M, Dixey J, Gillet P, Loeuille D, McGrath R, Reginato A, Shiang R, Wright G, Netter P, Williams C, Kingsley DM (2002) Mutations in ANKH cause chondrocalcinosis. Am J Hum Genet 71:933-40

Phillips MS, Lawrence R, Sachidanandam R, Morris AP, Balding DJ, Donaldson MA, Studebaker JF, et al. (2003) Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. Nat Genet 33:382-7

Ploski R, Flato B, Vinje O, Maksymowych W, Forre O, Thorsby E (1995) Association to HLA-DRB1*08, HLA-DPB1*0301 and homozygosity for an HLA-linked proteasome gene in juvenile ankylosing spondylitis. Hum Immunol 44:88-96

Ploski R, Maksymowych W, Forre O (1996) HLA-DR8 and susceptibility to acute anterior uveitis in ankylosing spondylitis: comment on the article by Monowarul Islam et al. Arthritis Rheum 39:351-2

Pollicino T, Pernice F, Campo S, Mesiti O, Misefari V, Pernice M, Raimondo G (1996) Severe outcome of hepatitis B virus (HBV) infection and lack of HBV e antigen-defective virus emergence in patients homozygous for HLA class I alleles. J Gen Virol 77 (Pt 8):1833-6

Poole AR (1998) The histopathology of ankylosing spondylitis: are there unifying hypotheses? Am J Med Sci 316:228-33

Poulsen L, Brosen K, Arendt-Nielsen L, Gram LF, Elbaek K, Sindrup SH (1996) Codeine and morphine in extensive and poor metabolizers of sparteine: pharmacokinetics, analgesic effect and side effects. Eur J Clin Pharmacol 51:289-95

Pritchard JK (2001) Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 69:124-37

Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65:220-8

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67:170-81

Pritchard JK, Donnelly P (2001) Case-control studies of association in structured or admixed populations. Theor Popul Biol 60:227-37
Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1-14
Pritchard JK, Cox NJ (2002) The allelic architecture of human disease genes: common disease-common variant...or not? Hum Mol Genet 11:2417-23
Prud'homme GJ, Piccirillo CA (2000) The inhibitory effects of transforming growth factor-beta-1 (TGF-beta1) in autoimmune diseases. J Autoimmun 14:23-42
Pulleyn LJ, Newton R, Adcock IM, Barnes PJ (2001) TGFbeta1 allele association with asthma severity. Hum Genet 109:623-7
Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 19:149-50
Queiro R, Torre JC, Gonzalez S, Lopez-Larrea C, Tinture T, Lopez-Lagunas I (2003) HLA Antigens May Influence the Age of Onset of Psoriasis and Psoriatic Arthritis. J Rheumatol 30:505-7
Ramos M, De Castro JA (2002) HLA-B27 and the pathogenesis of spondyloarthritis. Tissue Antigens 60:191-205
Ramos-Remus C, Russell AS, Gomez-Vargas A, Hernandez-Chavez A, Maksymowych WP, Gamez-Nava JI, Gonzalez-Lopez L, Garcia-Hernandez A, Meono-Morales E, Burgos-Vargas R, Suarez-Almazor ME (1998) Ossification of the posterior longitudinal ligament in three geographically and genetically different populations of ankylosing spondylitis and other spondyloarthropathies. Ann Rheum Dis 57:429-33
Randall LL, Hardy SJ (1989) Unity in function in the absence of consensus in sequence: role of leader peptides in export. Science 243:1156-9
Redpath S, Ghazal P, Gascoigne NR (2001) Hijacking and exploitation of IL-10 by intracellular pathogens. Trends Microbiol 9:86-92
Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. Nature 411:199-204
Reichenberger E, Tiziani V, Watanabe S, Park L, Ueki Y, Santanna C, Baur ST, Shiang R, Grange DK, Beighton P, Gardner J, Hamersma H, Sellars S, Ramesar R, Lidral AC, Sommer A, Raposo do Amaral CM, Gorlin RJ, Mulliken JB, Olsen BR (2001) Autosomal dominant craniometaphyseal dysplasia is caused by mutations in the transmembrane protein ANK. Am J Hum Genet 68:1321-6
Reveille JD (2002) Progress in whole genome scans: the North American experience. Clin Exp Rheumatol 20:590
Ricci-Vitiani L, Vacca A, Potolicchio I, Scarpa R, Bitti P, Sebastiani G, Passiu G, Mathieu A, Sorrentino R (2000) MICA gene triplet repeat polymorphism in patients with HLA-B27 positive and negative ankylosing spondylitis from Sardinia. J Rheumatol 27:2193-7
Richter-Hintz D, Their R, Steinwachs S, Kronenberg S, Fritsche E, Sachs B, Wulferink M, Tonn T, Esser C (2003) Allelic variants of drug metabolizing enzymes as risk factors in psoriasis. J Invest Dermatol 120:765-70
Ringrose JH (1999) HLA-B27 associated spondyloarthropathy, an autoimmune disease based on crossreactivity between bacteria and HLA-B27? Ann Rheum Dis 58:598-610
Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, et al. (1989) Identification of the cystic fibrosis

gene: cloning and characterization of complementary DNA. Science 245:1066-73

Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, McLeod RS, Griffiths AM, Green T, Brettin TS, Stone V, Bull SB, Bitton A, Williams CN, Greenberg GR, Cohen Z, Lander ES, Hudson TJ, Siminovitch KA (2000) Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. Am J Hum Genet 66:1863-70

Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, Delmonte T, et al. (2001) Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet 29:223-8

Risch N (1990a) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:222-8

Risch N (1990b) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 46:229-41

Risch N (1990c) Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. Am J Hum Genet 46:242-53

Risch NJ (2000) Searching for genetic determinants in the new millennium. Nature 405:847-56

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516-7

Risch N, Teng J (1998) The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. Genome Res 8:1273-88

Roberts AB, Anzano MA, Lamb LC, Smith JM, Sporn MB (1981) New class of transforming growth factors potentiated by epidermal growth factor: isolation from non-neoplastic tissues. Proc Natl Acad Sci U S A 78:5339-43

Robinson J, Waller MJ, Parham P, de Groot N, Bontrop R, Kennedy LJ, Stoehr P, Marsh SG (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res 31:311-4

Robinson WP, van der Linden SM, Khan MA, Rentsch HU, Cats A, Russell A, Thomson G (1989) HLA-Bw60 increases susceptibility to ankylosing spondylitis in HLA-B27+ patients. Arthritis Rheum 32:1135-41

Romano M, Sironi M, Toniatti C, Polentarutti N, Fruscella P, Ghezzi P, Faggioni R, Luini W, van Hinsbergh V, Sozzani S, Bussolino F, Poli V, Ciliberto G, Mantovani A (1997) Role of IL-6 and its soluble receptor in induction of chemokines and leukocyte recruitment. Immunity 6:315-25

Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, et al. (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 245:1059-65

Rosenbaum JT (1989) Characterization of uveitis associated with spondyloarthritis. J Rheumatol 16:792-6

Rosenblum BB, Lee LG, Spurgeon SL, Khan SH, Menchen SM, Heiner CR, Chen SM (1997) New dye-labeled terminators for improved DNA sequencing patterns. Nucleic Acids Res 25:4500-4

Rosenow E, Strimlan CV, Muhm JR, Ferguson RH (1977) Pleuropulmonary manifestations of ankylosing spondylitis. Mayo Clin Proc 52:641-9

Rostami-Hodjegan A, Lennard MS, Woods HF, Tucker GT (1998) Meta-analysis of studies of the CYP2D6 polymorphism in relation to lung cancer and Parkinson's disease. Pharmacogenetics 8:227-38

Rubin LA, Amos CI, Wade JA, Martin JR, Bale SJ, Little AH, Gladman DD, Bonney GE, Rubenstein JD, Siminovitch KA (1994) Investigating the genetic basis for ankylosing spondylitis. Linkage studies with the major histocompatibility complex region. Arthritis Rheum 37:1212-20

Rudwaleit M, Siegert S, Yin Z, Eick J, Thiel A, Radbruch A, Sieper J, Braun J (2001) Low T cell production of TNFalpha and IFNgamma in ankylosing spondylitis: its relation to HLA-B27 and influence of the TNF-308 gene polymorphism. Ann Rheum Dis 60:36-42

Saari M, Vuorre I, Kaila J, Lahti R (1978) Family studies of ocular manifestations in arthritis. Can J Ophthalmol 13:144-51

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928-33

Saharinen J, Taipale J, Monni O, Keski-Oja J (1998) Identification and characterization of a new latent transforming growth factor-beta-binding protein, LTBP-4. J Biol Chem 273:18459-69

Said-Nahal R, Miceli-Richard C, Gautreau C, Tamouza R, Borot N, Porcher R, Charron D, Dougados M, Breban M (2002) The role of HLA genes in familial spondyloarthropathy: a comprehensive study of 70 multiplex families. Ann Rheum Dis 61:201-6

Sambrook J, Russell DW (2001) Molecular cloning, a laboratory manual. Vol 2. Cold Spring Harbor Laboratory Press, New York

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463-7

Sanmarti R, Ercilla MG, Brancos MA, Cid MC, Collado A, Rotes-Querol J (1987) HLA class II antigens (DR, DQ loci) and peripheral arthritis in ankylosing spondylitis. Ann Rheum Dis 46:497-500

Saraux A, de Saint-Pierre V, Baron D, Valls I, Koreichi D, Youinou P, Le Goff P (1995) The HLA B27 antigen-spondylarthropathy association. Impact on clinical expression. Rev Rhum Engl Ed 62:487-91

Sargent CA, Dunham I, Trowsdale J, Campbell RD (1989) Human major histocompatibility complex contains genes for the major heat shock protein HSP70. Proc Natl Acad Sci U S A 86:1968-72

Sasaki T, Tahira T, Suzuki A, Higasa K, Kukita Y, Baba S, Hayashi K (2001) Precise estimation of allele frequencies of single-nucleotide polymorphisms by a quantitative SSCP analysis of pooled DNA. Am J Hum Genet 68:214-8

Sayer D, Whidborne R, Brestovac B, Trimboli F, Witt C, Christiansen F (2001) HLA-DRB1 DNA sequencing based typing: an approach suitable for high throughput typing including unrelated bone marrow registry donors. Tissue Antigens 57:46-54

Schlosstein L, Terasaki PI, Bluestone R, Pearson CM (1973) High association of an HL-A antigen, W27, with ankylosing spondylitis. N Engl J Med 288:704-6

Schrijver HM, Crusius JB, Uitdehaag BM, Garcia Gonzalez MA, Kostense PJ, Polman CH, Pena AS (1999) Association of interleukin-1beta and interleukin-1 receptor antagonist genes with disease severity in MS. Neurology 52:595-9

Schulte CM, Goebell H, Roher HD, Schulte KM (2001) C-509T polymorphism in the TGFB1 gene promoter: impact on Crohn's disease susceptibility and clinical course? Immunogenetics 53:178-82

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W (2000) PipMaker--a web server for aligning two genomic DNA sequences. Genome Res 10:577-86

Sham PC, Purcell S, Cherny SS, Abecasis GR (2002) Powerful regression-based quantitative-trait linkage analysis of general pedigrees. Am J Hum Genet 71:238-53

Shankarkumar U, Ghosh K, Mohanty D (2002) HLA B27 polymorphism in Western India. Tissue Antigens 60:98-101

Shifman S, Darvasi A (2001) The value of isolated populations. Nat Genet 28:309-10

Shull MM, Ormsby I, Kier AB, Pawlowski S, Diebold RJ, Yin M, Allen R, Sidman C, Proetzel G, Calvin D, et al. (1992) Targeted disruption of the mouse transforming growth factor-beta 1 gene results in multifocal inflammatory disease. Nature 359:693-9

Sims AM, Barnardo M, Broxholme J, Herzberg I, Bradbury L, Calin A, Wordsworth BP, Darke C, Brown MA (2004) Mapping the non-B27 MHC Associations of Ankylosing Spondylitis (AS). Rheumatology submitted

Singer VL, Jones LJ, Yue ST, Haugland RP (1997) Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. Anal Biochem 249:228-38

Siren MK, Sareneva H, Lokki ML, Koskimies S (1996) Unique HLA antigen frequencies in the Finnish population. Tissue Antigens 48:703-7

Skarsvag S, Hansen KE, Holst A, Moen T (1992) Distribution of HLA class II alleles among Scandinavian patients with systemic lupus erythematosus (SLE): an increased risk of SLE among non[DRB1*03,DQA1*0501,DQB1*0201] class II homozygotes? Tissue Antigens 40:128-33

Slateva K, Elsner HA, Albis-Camps M, Blasczyk R (2001) HLA-DRB fluorotyping by dark quenching and automated analysis. Tissue Antigens 58:250-4

Smith LM, Fung S, Hunkapiller MW, Hunkapiller TJ, Hood LE (1985) The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. Nucleic Acids Res 13:2399-412

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE (1986) Fluorescence detection in automated DNA sequence analysis. Nature 321:674-9

Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 58:1323-37

Sonel B, Tutkak H, Duzgun N (2002) Serum levels of IL-1 beta, TNF-alpha, IL-8, and acute phase proteins in seronegative spondyloarthropathies. Joint Bone Spine 69:463-7

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506-16

Spielman RS, Ewens WJ (1998) A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet 62:450-8

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978-89

Stokkers PC, Reitsma PH, Tytgat GN, van Deventer SJ (1999) HLA-DR and -DQ phenotypes in inflammatory bowel disease: a meta-analysis. Gut 45:395-401

Stuven T, Griese EU, Kroemer HK, Eichelbaum M, Zanger UM (1996) Rapid detection of CYP2D6 null alleles by long distance- and multiplex-polymerase chain reaction. Pharmacogenetics 6:417-21

Suarez-Almazor ME, Russell AS (1987) B27 homozygosity and ankylosing spondylitis. J Rheumatol 14:302-4

Sugiura Y, Niimi T, Sato S, Yoshinouchi T, Banno S, Naniwa T, Maeda H, Shimizu S, Ueda R (2002) Transforming growth factor beta1 gene polymorphism in rheumatoid arthritis. Ann Rheum Dis 61:826-8

Svejgaard A, Ryder LP (1994) HLA and disease associations: detecting the strongest association. Tissue Antigens 43:18-27

Syrris P, Carter ND, Metcalfe JC, Kemp PR, Grainger DJ, Kaski JC, Crossman DC, Francis SE, Gunn J, Jeffery S, Heathcote K (1998) Transforming growth factor-beta1 gene polymorphisms and coronary artery disease. Clin Sci (Lond) 95:659-67

Syvanen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2:930-42

Tabor HK, Risch NJ, Myers RM (2002) Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 3:391-7

Taillon-Miller P, Bauer-Sardina I, Saccone NL, Putzel J, Laitinen T, Cao A, Kere J, Pilia G, Rice JP, Kwok PY (2000) Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. Nat Genet 25:324-8

Tamouza R, Mansour I, Bouguacha N, Klayme S, Djouadi K, Laoussadi S, Azoury M, Dulphy N, Ramasawmy R, Krishnamoorthy R, Toubert A, Naman R, Charron D (2001) A new HLA-B*27 allele (B*2719) identified in a Lebanese patient affected with ankylosing spondylitis. Tissue Antigens 58:30-3

Taurog JD, Lowen L, Forman J, Hammer RE (1988) HLA-B27 in inbred and non-inbred transgenic mice. Cell surface expression and recognition as an alloantigen in the absence of human beta 2-microglobulin. J Immunol 141:4020-3

Taurog JD, Maika SD, Simmons WA, Breban M, Hammer RE (1993) Susceptibility to inflammatory disease in HLA-B27 transgenic rat lines correlates with the level of B27 expression. J Immunol 150:4168-78

Taurog JD, Richardson JA, Croft JT, Simmons WA, Zhou M, Fernandez-Sueiro JL, Balish E, Hammer RE (1994) The germfree state prevents development of gut and joint inflammatory disease in HLA-B27 transgenic rats. J Exp Med 180:2359-64

Terwilliger JD, Ott J (1992) A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. Hum Hered 42:337-46

Thomson G (1995) Mapping disease genes: family-based association studies. Am J Hum Genet 57:487-98

Timms AE, Zhang Y, Bradbury L, Wordsworth BP, Brown MA (2003) The role of ANKH in Ankylosing Spondylitis. Arthritis Rheum in press

Tiret L, Poirier O, Nicaud V, Barbaux S, Herrmann SM, Perret C, Raoux S, Francomme C, Lebard G, Tregouet D, Cambien F (2002) Heterogeneity of linkage disequilibrium in human genes has implications for association studies of common diseases. Hum Mol Genet 11:419-29

Tishkoff SA, Pakstis AJ, Ruano G, Kidd KK (2000) The accuracy of statistical methods for estimation of haplotype frequencies: an example from the CD4 locus. Am J Hum Genet 67:518-22

Tomfohrde J, Silverman A, Barnes R, Fernandez-Vina MA, Young M, Lory D, Morris L, Wuepper KD, Stastny P, Menter A, et al. (1994) Gene for familial psoriasis susceptibility mapped to the distal end of human chromosome 17q. Science 264:1141-5

Tremmel M, Opelz G, Mytilineos J (1999) High-resolution typing for HLA-DRB1*15 amd -DRB1*16 by fluorescence-marked sequence-specific priming (TaqMan assay). Tissue Antigens 54:508-16

Tsuchiya N, Shiota M, Moriyama S, Ogawa A, Komatsu-Wakui M, Mitsui H, Geraghty DE, Tokunaga K (1998) MICA allele typing of HLA-B27 positive Japanese patients with seronegative spondylarthropathies and healthy individuals: differential linkage disequilibrium with HLA-B27 subtypes. Arthritis Rheum 41:68-73

Tsui FW, Tsui HW, Cheng EY, Stone M, Payne U, Reveille JD, Shulman MJ, Paterson AD, Inman RD (2003) Novel genetic markers in the 5'-flanking region of ANKH are associated with ankylosing spondylitis. Arthritis Rheum 48:791-7

Tuomilehto-Wolf E, Tuomilehto J, Cepaitis Z, Lounamaa R (1989) New susceptibility haplotype for type 1 diabetes. DIME Study Group. Lancet 2:299-302

Uryu N, Maeda M, Ota M, Tsuji K, Inoko H (1990) A simple and rapid method for HLA-DRB and -DQB typing by digestion of PCR-amplified DNA with allele specific restriction endonucleases. Tissue Antigens 35:20-31

van der Linden S, Valkenburg HA, Cats A (1984a) Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum 27:361-8

van der Linden SM, Valkenburg HA, de Jongh BM, Cats A (1984b) The risk of developing ankylosing spondylitis in HLA-B27 positive individuals. A comparison of relatives of spondylitis patients with the general population. Arthritis Rheum 27:241-9

van der Paardt M, Crusius JB, Garcia-Gonzalez MA, Baudoin P, Kostense PJ, Alizadeh BZ, Dijkmans BA, Pena AS, van der Horst-Bruinsma IE (2002a) Interleukin-1beta and interleukin-1 receptor antagonist gene polymorphisms in ankylosing spondylitis. Rheumatology (Oxford) 41:1419-23

van der Paardt M, Dijkmans B, Giltay E, van der Horst-Bruinsma I (2002b) Dutch patients with familial and sporadic ankylosing spondylitis do not differ in disease phenotype. J Rheumatol 29:2583-4

Van Eerdewegh P, Little RD, Dupuis J, Del Mastro RG, Falls K, Simon J, Torrey D, et al. (2002) Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 418:426-30

Vargas-Alarcon G, Londono JD, Hernandez-Pacheco G, Gamboa R, Castillo E, Pacheco-Tena C, Cardiel MH, Granados J, Burgos-Vargas R (2002a) Heat shock protein 70 gene polymorphisms in Mexican patients with spondyloarthropathies. Ann Rheum Dis 61:48-51

Vargas-Alarcon G, Londono JD, Hernandez-Pacheco G, Pacheco-Tena C, Castillo E, Cardiel MH, Granados J, Burgos-Vargas R (2002b) Effect of HLA-B and HLA-DR genes on susceptibility to and severity of spondyloarthropathies in Mexican patients. Ann Rheum Dis 61:714-7

Varilo T, Savukoski M, Norio R, Santavuori P, Peltonen L, Jarvela I (1996) The age of human mutation: genealogical and linkage disequilibrium analysis of the CLN5 mutation in the Finnish population. Am J Hum Genet 58:506-12

Varilo T, Paunio T, Parker A, Perola M, Meyer J, Terwilliger JD, Peltonen L (2003) The interval of linkage disequilibrium (LD) detected with microsatellite and

SNP markers in chromosomes of Finnish populations with different histories. Hum Mol Genet 12:51-9
Verner K, Schatz G (1988) Protein translocation across membranes. Science 241:1307-13
Veys EM, van Leare M (1973) Serum IgG, IgM, and IgA levels in ankylosing spondylitis. Ann Rheum Dis 32:493-6
Voss H, Schwager C, Wirkner U, Sproat B, Zimmermann J, Rosenthal A, Erfle H, Stegemann J, Ansorge W (1989) Direct genomic fluorescent on-line sequencing and analysis using in vitro amplification of DNA. Nucleic Acids Res 17:2517-27
Wahl SM (1999) TGF-beta in the evolution and resolution of inflammatory and immune processes. Introduction. Microbes Infect 1:1247-9
Walsh EC, Mather KA, Schaffner SF, Farwell L, Daly MJ, Patterson N, Cullen M, Carrington M, Bugawan TL, Erlich H, Campbell J, Barrett J, Miller K, Thomson G, Lander ES, Rioux JD (2003) An integrated haplotype map of the human major histocompatibility complex. Am J Hum Genet 73:580-90
Wanstrat A, Wakeland E (2001) The genetics of complex autoimmune diseases: non-MHC susceptibility genes. Nat Immunol 2:802-9
Watanabe RM, Ghosh S, Langefeld CD, Valle TT, Hauser ER, Magnuson VL, Mohlke KL, et al. (2000) The Finland-United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. II. An autosomal genome scan for diabetes-related quantitative-trait loci. Am J Hum Genet 67:1186-200
Weedon MN, Schwarz PE, Horikawa Y, Iwasaki N, Illig T, Holle R, Rathmann W, Selisko T, Schulze J, Owen KR, Evans J, Del Bosque-Plata L, Hitman G, Walker M, Levy JC, Sampson M, Bell GI, McCarthy MI, Hattersley AT, Frayling TM (2003) Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility. Am J Hum Genet 73:1208-12
Weinshilboum R (2003) Inheritance and drug response. N Engl J Med 348:529-37
Werner M, Sych M, Herbon N, Illig T, Konig IR, Wjst M (2002) Large-scale determination of SNP allele frequencies in DNA pools using MALDI-TOF mass spectrometry. Hum Mutat 20:57-64
Westendorp RG, Langermans JA, Huizinga TW, Elouali AH, Verweij CL, Boomsma DI, Vandenbroucke JP, Vandenbrouke JP (1997) Genetic influence on cytokine production and fatal meningococcal disease. Lancet 349:170-3
Westman P, Partanen J, Leirisalo-Repo M, Koskimies S (1994a) Different DRB1*04 alleles predominate in the Finnish random population and in HLA-B27-positive subpopulations. Tissue Antigens 44:329-31
Westman P, Partanen J, Leirisalo-Repo M, Koskimies S (1994b) HSP70-Hom NcoI polymorphism and HLA-associations in the Finnish population and in patients with ankylosing spondylitis or reactive arthritis. Eur J Immunogenet 21:81-90
Westman P, Partanen J, Leirisalo-Repo M, Koskimies S (1995) TAP1 and TAP2 polymorphism in HLA-B27-positive subpopulations: no allelic differences in ankylosing spondylitis and reactive arthritis. Hum Immunol 44:236-42
Westman P, Leirisalo-Repo M, Partanen J, Koskimies S (1996) A comparative study of HLA genes in HLA-B27 positive ankylosing spondylitis and HLA-B27 positive peripheral reactive arthritis. Arthritis Rheum 39:943-9
Will R, Palmer R, Bhalla AK, Ring F, Calin A (1989) Osteoporosis in early ankylosing spondylitis: a primary pathological event? Lancet 2:1483-5

Williams CJ, Zhang Y, Timms A, Bonavita G, Caeiro F, Broxholme J, Cuthbertson J, Jones Y, Marchegiani R, Reginato A, Russell RG, Wordsworth BP, Carr AJ. Brown MA (2002) Autosomal dominant familial calcium pyrophosphate dihydrate deposition disease is caused by mutation in the transmembrane protein ANKH. Am J Hum Genet 71:985-91

Wordsworth BP, Allsopp CE, Young RP, Bell JI (1990) HLA-DR typing using DNA amplification by the polymerase chain reaction and sequential hybridization to sequence-specific oligonucleotide probes. Immunogenetics 32:413-8

Yabuki K, Ota M, Goto K, Kimura T, Nomura E, Ohno S, Mizuki N, Katsuyama Y. Makysymowych WP, Bahram S, Kimura M, Inoko H (1999) Triplet repeat polymorphism in the MICA gene in HLA-B27 positive and negative caucasian patients with ankylosing spondylitis. Hum Immunol 60:83-6

Yamada Y, Miyauchi A, Goto J, Takagi Y, Okuizumi H, Kanematsu M, Hase M, Takai H, Harada A, Ikeda K (1998) Association of a polymorphism of the transforming growth factor-beta1 gene with genetic susceptibility to osteoporosis in postmenopausal Japanese women. J Bone Miner Res 13:1569-76

Yamada Y, Okuizumi H, Miyauchi A, Takagi Y, Ikeda K, Harada A (2000) Association of transforming growth factor beta1 genotype with spinal osteophytosis in Japanese women. Arthritis Rheum 43:452-60

Yamaguchi A, Tsuchiya N, Mitsui H, Shiota M, Ogawa A, Tokunaga K, Yoshinoya S, Juji T, Ito K (1995) Association of HLA-B39 with HLA-B27-negative ankylosing spondylitis and pauciarticular juvenile rheumatoid arthritis in Japanese patients. Evidence for a role of the peptide-anchoring B pocket. Arthritis Rheum 38:1672-7

Yi AK, Tuetken R, Redford T, Waldschmidt M, Kirsch J, Krieg AM (1998) CpG motifs in bacterial DNA activate leukocytes through the pH-dependent generation of reactive oxygen species. J Immunol 160:4755-61

Zippert R, Bassler A, Holmer SR, Hengstenberg C, Schunkert H (2000) Eleven single nucleotide polymorphisms and one triple nucleotide insertion of the human TGF-beta III receptor gene. J Hum Genet 45:250-3

Zou J, Zhang Y, Thiel A, Rudwaleit M, Shi SL, Radbruch A, Poole R, Braun J, Sieper J (2003) Predominant cellular immune response to the cartilage autoantigenic G1 aggrecan in ankylosing spondylitis and rheumatoid arthritis. Rheumatology (Oxford)

Zwirner NW, Dole K, Stastny P (1999) Differential surface expression of MICA by endothelial cells, fibroblasts, keratinocytes, and monocytes. Hum Immunol 60:323-30

APPENDIX 1

Questionnaire for Ankylosing Spondylitis / Selkärankareumakysely

MAIN

Variable	Format	Kuvaus	Description
RECID**	integer	*Record ID, for internal use*	*Record ID, for internal use*
CENTERID	char(2)	*Center ID*: 01-99	*Center ID*: 01-99
FAMILYID	char(4)	*Family ID*: 0001-9999	*Family ID*: 0001-9999
INDIVID	char(3)	*Individual ID:* 000= proband 001-999= other relatives	*Individual ID:* 000= proband 001-999= other relatives
PTID**	char(9)	*for internal use*	*for internal use*
SSN**	char(11)	*11-character Social Security number*	*11-character Social Security number*
BDATE	char(6)	*Date of birth:* ddmmyyyy	*Date of birth:* ddmmyyyy
L_NAME**	char(30)	*Last name*	*Last name*
F_NAME**	char(50)	*First name(s)*	*First name(s)*
ADDRESS**	char(50)	*Address, street*	*Address, street*
CITY**	char(30)	*Address, city*	*Address, city*
ZIPCODE**	char(6)	*Address, postal code*	*Address, postal code*
TEL_1**	char(20)	*Telephone - 1*	*Telephone - 1*
TEL_2**	char(20)	*Telephone - 2*	*Telephone - 2*
TEL_3**	char(20)	*Telephone - 3*	*Telephone - 3*
Q1	char(1)	***K1**. Mikä on koulutuksenne? ..* 1= kansankoulu 2= ammattikoulu 3= opistotutkinto 4= korkeakoulu 9= don't know	*2. What is your education? ..* 1= kansankoulu 2= ammattikoulu 3= opistotutkinto 4= korkeakoulu 9= don't know
Q2YEAR	char(4)	***K2**. Milloin selkärankareumanne oireet alkovait? Vuonna:* 9999= don't know	***1.a**. When did your ankylosing spondylitis (AS) begin? year:* 9999= don't know
Q2AGE	char(2)	***K2**. ... ikäisenä* 99= don't know	***1.a.i**. age:* 99= don't know
Q3YEAR	char(4)	***K3**. Milloin selkärankareumanne*	***1.b**. When was the*

		oireet alkovait? Vuonna: 9999= don't know	*diagnosis of AS made? year:* 9999= don't know
Q3AGE	char(2)	***K3***. *... ikäisenä* 99= don't know	***1.b.i***. *age:* 99= don't know
Q4	char(1)	***K4***. *Oletteko ollut reumatautien...:* 1= yes 2= no 9= don't know	***1.c***. *Have you ever seen a rheumatologist:* 1= yes 2= no 9= don't know
Q5	char(2)	***K5***. *Mikäli vastasitte KYLLÄ, ... vastaanotolla selkärankareumanne takia? kertaa* 99= don't know	***1.c.i***. *If YES, on average, how many visits do you make to your rheumatologist per year because of your AS?* 99= don't know
Q6	char(1)	***K6***. *Onko selästänne koskaan ptettu röntgenkuvaa?* 1= yes 2= no 9= don't know	***1.d***. *Have you ever had X-rays taken of your back?* 1= yes 2= no 9= don't know
Q7DATE	char(10) dd.mm.yyyy	***K7***. *Mikäli vastasitte KYLLÄ, niin mainitkaa aika ja paika missä viimesin...*	***1.e***. *If YES, please say when and where the most recent ones were taken. Year:*
Q7PLACE**	char(100)	***K7***. *paika*	***1.e.i***. *Where (place):*
Q8	char(1)	***K8***. *Oletteko tällä hetkellä työssä? ...* 1= yes 2= no 9= don't know	***3.a.i***. *Are you currently employed?* 1= yes 2= no 9= don't know
Q9	char(1)	***K9***. *Mikä vastauksenne on KYLLÄ, niin kuinka pitkään...* 1= en ollenkaan 2= vähemmän kuin 1 viikko 3= 1-4 viikkoa 4= 1-3 kuukautta 5= yli 3 kuukautta 9= don't know	***3.b***. *If YES, how much time off work have you had per year because of your AS:* 1= none 2= less than 1 week 3= 1-4 weeks 4= 1-3 months 5= more than 3 months 9= don't know
Q10	char(1)	***K10***. *Mikäli vastauksenne oi EI, niin oletteko* 1= työtön / hakemassa työtä 2= eläkkeellä 9= don't know	***3.c***. *If NO, are you:* 1= unemployed & intending to work 2= retired 9= don't know
Q11	char(2)	***K11***. *Jos olette... vuotiaana*	***3.d.i***. *If RETIRED, at*

		99= don't know	*what age did you retire?* 99= don't know
Q12	char(2)	***K12***. *Minkä ikäisenä ... vuotiaana* 99= don't know	***3.d.i.i***. *What is the usual age of retirement for your occupation?* 99= don't know
Q13	char(1)	***K13***. *Oletteko työttömänä - eläkkeellä:* 1= pelkästään selkärankareumanne takia 2= osittain selkärankareumanne takia 3= muun kuin selkärankareumanne takia 9= don't know	***3.e***. *Are you UNEMPLOYED / RETIRED:* 1= Solely because of your AS? 2= Partly because of your AS? 3= For reasons other than your AS? 9= don't know
		K14. Mitkä ovat mielestänne PÄÄSIALLISET selkärankareumoireenne?	**4. In your opinion, which is the main symptom of your AS? (please tick one)**
Q14_1	char(1)	***K14.1***. *väsymys* 0= no 1= yes	***4.1***. *Fatigue:* 0= no 1= yes
Q14_2	char(1)	***K14.2***. *selkäkipu (niska / selkä / lonkka)*K14. Mitkä ovat mielestänne PÄÄSIALLISET selkärankareumoireenne? 0= no 1= yes	***4.2***. *Spinal pain (neck / back / hip):* 0= no 1= yes
Q14_3	char(1)	***K14.3***. *jäykkyys* 0= no 1= yes	***4.3***. *Stiffness:* 0= no 1= yes
Q14_4	char(1)	***K14.4***. *nivelkipu / turvotus* 0= no 1= yes	***4.4***. *Joint pain / swelling:* 0= no 1= yes
Q14_5	char(1)	***K14.5***. *kosketus / paineluarat alueet* 0= no 1= yes	***4.5***. *Areas tender to touch / pressure:* 0= no 1= yes
Q14_6	char(1)	***K14.6***. *ei pysty erittelemään* 0= no 1= yes	***4.6***. *Cannot distinguish one main symptom:* 0= no 1= yes
Q14_7	char(1)	***K14.7a***. *muu, tarkentakaa* 0= no 1= yes	***4.7a***. *Other (please specify:)* 0= no 1= yes

Q14_7TXT	char(50)	***K14.7b***. *tarkentakaa*	***4.7b***. *Specification*
Q15	char(1)	***K15***. *Onko Teillä ollut värikalvontulehdus..* 1= yes 2= no **(go to K23)** 9= don't know	***5.a***. *Do you get iritis (uveitis)?* 1= yes 2= no **(go to 6.a)** 9= don't know
Q16	char(1)	***K16***. *Mikäli vastasitte KYLLÄ, kuka sen totesi?* 1= yleislääkäri 2= reumatautien erikoislääkäri 3= silmälääkäri 4= joku muu kuin lääkäri 8= irrelevant (Q15=2) 9= don't know	***5.a.i***. *If YES, who made the diagnosis?* 1= GP 2= Rheumatologist 3= Eye specialist 4= No doctor 8= irrelevant (**5.a**=2) 9= don't know
Q17	char(1)	***K17***. *Onko Teillä ollut silmän värikalvontulehdukseen...* 1= yes 2= no 8= irrelevant (Q15=2) 9= don't know	***5.b***. *Have you had any treatment for it?* 1= yes 2= no 8= irrelevant (**5.a**=2) 9= don't know
Q18	char(200)	***K18***. *Mikäli vastasitte KYLLÄ, niin mainitkaa tarkemmin minkälaisen hoidon olette saanut*	***5.b.i***. *If YES, please specify*
Q19	char(2)	***K19***. *Kuinka monta kertaa Teillä on ollut silmän värikalvontulehdus? ... kertaa* 88= irrelevant (Q15=2) 99= missing	***5.c***. *Approximatelly how many attacks have you had overall?* 88= irrelevant (**5.a**=2) 99= missing
Q20	char(2)	***K20***. *Kuinka monta kertaa Teillä on yleensä silmän värikalvontulehdus... kertaa* 88= irrelevant (Q15=2) 99= missing	***5.d***. *Approximatelly how many attacks do you have per year?* 88= irrelevant (**5.a**=2) 99= missing
Q21DATE	char(10) dd.mm.yyyy	***K21***. *Milloin Teillä oli silmän värikalvontulehdus viimeksi?* 88.88.8888= irrelevant (Q15=2) 99.99.9999= missing	***5.e***. *When was your last attack?* 88.88.8888= irrelevant (**5.a**=2) 99.99.9999= missing
Q22	char(1)	***K22***. *Onko silmän värikalvontulehdus aiheuttanut pysyvän...* 1= yes 2= no 8= irrelevant (Q15=2) 9= don't know	***5.f***. *Has the uveitis resulted in persistent deterioration of vision* 1= yes 2= no 8= irrelevant (**5.a**=2) 9= don't know
Q23	char(1)	***K23***. *Onko Teillä psoriasi (ihottuma)?* 1= yes	***6.a***. *Do you get psoriasis?* 1= yes

		2= no **(go to K27)** 9= don't know	2= no **(go to 7.a)** 9= don't know
Q24	char(1)	***K24***. *Mikäli vastasitte KYLLÄ, niin mainitkaa kuka teki ..* 1= yleislääkäri 2= reumatautien erikoislääkäri 3= ihotautilääkäri 4= joku muu kuin lääkäri 8= irrelevant (Q23=2) 9= don't know	***6.a.i***. *If YES, who made the diagnosis?* 1= GP 2= Rheumatologist 3= Eye specialist 4= No doctor 8= irrelevant (**6.a**=2) 9= don't know
Q25	char(1)	***K25***. *Onko Teillä ollut jokin hoito psoriaasikseene?* 1= yes 2= no 8= irrelevant (Q23=2) 9= don't know	***6.b***. *Have you had any tratment for it?* 1= yes 2= no 8= irrelevant (**6.a**=2) 9= don't know
Q26	char(200)	***K26***. *Mikäli vastasitte KYLLÄ, niin mainitkaa tarkemmin hoidosta?*	***6.b.i***. *If YES, please specify*
Q27	char(1)	***K27***. *Onko Teillä todettu Crohnin sairaus?* 1= yes 2= no 9= don't know	***7.a***. *Have you ever been diagnosed as having Crohn's disease?* 1= yes 2= no 9= don't know
Q28	char(1)	***K28***. *Mikäli vastasitte KYLLÄ, kuka teki diagnoosin?* 1= yleislääkäri 2= reumatautien erikoislääkäri 3= vatsatautien erikoislääkäri 4= joku muu kuin lääkäri 9= don't know	***7.a.i***. *If YES, who made the diagnosis?* 1= GP 2= Rheumatologist 3= Eye specialist 4= No doctor 9= don't know
Q29	char(1)	***K29***. *Onko Teillä todettu haavainen paksusuolen tulehdus?* 1= yes 2= no 9= don't know	***7.b***. *Have you ever been diagnosed as having Ulcerative colitis?* 1= yes 2= no 9= don't know
Q30	char(1)	***K30***. *Mikäli vastasitte KYLLÄ, kuka teki diagnoosin?* 1= yleislääkäri 2= reumatautien erikoislääkäri 3= vatsatautien erikoislääkäri 4= joku muu kuin lääkäri 9= don't know	***7.b.i***. *If YES, who made the diagnosis?* 1= GP 2= Rheumatologist 3= Eye specialist 4= No doctor 9= don't know
		K31. Mitkä ovat mielestänne PÄÄSIALLISET selkärankareumoireenne?	**8. Have you had during past year unexplained (i.e. not due to an**

			accident or injury) pain or swelling on any of the following joints?
Q31_1	char(1)	***K31.1***. *käsi* 0= no 1= yes	***8.1***. *Hand* 0= no 1= yes
Q31_2	char(1)	***K31.2***. *kyynärpää* 0= no 1= yes	***8.2***. *Elbow* 0= no 1= yes
Q31_3	char(1)	***K31.3***. *olkapää* 0= no 1= yes	***8.3***. *Shoulder* 0= no 1= yes
Q31_4	char(1)	***K31.4***. *lonkka* 0= no 1= yes	***8.4***. *HIP* 0= no 1= yes
Q31_5	char(1)	***K31.5***. *polvi* 0= no 1= yes	***8.5***. *Knee* 0= no 1= yes
Q31_6	char(1)	***K31.6***. *nilkka* 0= no 1= yes	***8.6***. *Ankle* 0= no 1= yes
Q31_7	char(1)	***K31.7***. *jalkaterä* 0= no 1= yes	***8.7***. *Foot* 0= no 1= yes
Q31_8	char(1)	***K31.8***. *kantapää* 0= no 1= yes	***8.8***. *Heel* 0= no 1= yes
Q32	char(1)	***K32***. *Onko Teillä koskaan tehty leikkausta niveliinne tai...* 1= yes 2= no 9= don't know	***9.a***. *Have you ever had surgery for your joint or spine as a result of your AS?* 1= yes 2= no 9= don't know Note: information on surgeries is table **surgeries**
Q34	char(1)	***K34***. *Kuinka mointa tuntia keskimäärin viikossa osallistutte urheiluun...* 1= 10 tai useampia tunteja 2= 5-9 tuntia 3= 2-4 tuntia 4= 1 tunti 5= 0 tuntia 9= don't know	***10.a***. *For how many hours per week on an average have you taken part in* sports, AS exercises or hydrotherapy during the last three months? 1= 10 and more hours 2= 5-9 hours 3= 2-4 hours 4= 1 hour

			5= 0 hours 9= don't know
Q35	char(2)	***K35***. ... *joka liikuntaharjoituksillanne on oireittenne lievitykseen* 0= Ei mitään 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Hyvin tehokas 99= don't know	***10.b***. *Place a vertical mark on the scale below to indicate the effectiveness which your exercise has on relieving your symptoms:* 0= None 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Very effective 99= don't know
Q36	char(2)	***K36***. ... *sairaudellanne on ollut vointiinne VIIMEISEN VIIKON AIKANA* 0= Ei mitään 1= 1 hour 2= 2 hours 3= 3 hours 4= 4 hours 5= 5 hours 6= 6 hours 7= 7 hours 8= 8 hours 9= 9 hours 10= 10 hours 11= Huonoin mahdollinen 99= don't know	***11.a***. *Place a vertical mark on the scale below to indicate the effect your disease has had on your well-being over the* ***last week****:* 0= None 1= 1 hour 2= 2 hours 3= 3 hours 4= 4 hours 5= 5 hours 6= 6 hours 7= 7 hours 8= 8 hours 9= 9 hours 10= 10 hours 11= Worst possible 99= don't know
Q37	char(2)	***K37***. ... *sairaudellanne on ollut vointiinne VIIMEISEN KUUDEN (6) kuukauden aikana* 0= Ei mitään 1= 1 hour 2= 2 hours 3= 3 hours 4= 4 hours 5= 5 hours	***11.b***. *Place a vertical mark on the scale below to indicate the effect your disease has had on your well-being over the* ***six months:*** 0= None 1= 1 hour 2= 2 hours

		6= 6 hours 7= 7 hours 8= 8 hours 9= 9 hours 10= 10 hours 11= Huonoin mahdollinen 99= don't know	3= 3 hours 4= 4 hours 5= 5 hours 6= 6 hours 7= 7 hours 8= 8 hours 9= 9 hours 10= 10 hours 11= Worst possible 99= don't know
Q38	char(1)	***K38****. ... tällä hetkellä jokin lääkitys selkärankareuman takia?* 1= yes 2= no 9= don't know	***12.a****. Are you currently taken any medication for your AS?* 1= yes 2= no 9= don't know Note: information on medication is given in table **medications**
			13. Please place a mark on each line below to indicate your answer to each question, relating to the PAST WEEK
Q41	char(2)	***K41****. Kuinka arvioitte kokemaanne uupumuksen / väsymyksen kokonaistasoa?* 0= Ei vaikutusta 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Hyvin vaikea 99= don't know	***13.1****. How you would describe the overall level of fatigue / tiredness you have experienced?* 0= None 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Very severe 99= don't know
Q42	char(2)	***K42****. Kuinka arvioitte selkärankareumastanne aiheutuvan niska-, selkä- tai ...* 0= Ei ollenkaan 1= 1 2= 2 3= 3 4= 4	***13.2****. How you would describe the overall level of AS* ***neck, back or hip*** *pain you have had?* 0= None 1= 1 2= 2 3= 3

		5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Hyvin paljon 99= don't know	4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Very severe 99= don't know
Q43	char(2)	***K43**.Onko muualla kuin niskassa, selässä tai lonkassa oleva nivelten kipu\turvotus...* 0= Ei ollenkaan 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Hyvin paljon 99= don't know	***13.3**. How you would describe the overall level of pain / swelling in joints other than **neck, back or hips** you have had?* 0= None 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Very severe 99= don't know
Q44	char(2)	***K44**. Kuinka arvioitte millä tahansa kehon alueella olevan kosketus- tai ...* 0= Ei vaikutusta 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Hyvin paljon 99= don't know	***13.4**. How you would describe the overall level of discomfort you have had from any areas tender to touch or preassure?* 0= None 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Very severe 99= don't know
Q45	char(2)	***K45**. Kuinka arvioitte selän aamujäykkyyden vaikuttaneen*	***13.5**. How you would describe the overall level*

		vointiinne päivän aikana? 0= Ei vaikutusta 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Hyvin paljon 99= don't know	*of morning stiffness you have had from the time you wake up?* 0= None 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Very severe 99= don't know
Q46	char(1)	***K46**. Kuinka kauan selkänne aamujäykkyys kestää?* 1= 0 min. 2= 10 min. 3= 20 min. 4= 30 min. 5= 40 min. 6= 50 min. 7= >= 1 tai useampi tunti 9= don't know	***13.6**. How long does your morning stiffness last from the time you wake up?* 1= 0 min. 2= 10 min. 3= 20 min. 4= 30 min. 5= 40 min. 6= 50 min. 7= >= 1 or more hours 9= don't know
			14. Please place a mark on each line below to indicate your level of ability with each of the following activities during the PAST WEEK
Q47	char(2)	***K47**. Sukkien tai sukkahousujen pukeminen päälle ilman apua tai apuvälineitä?* 0= Helppoa 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10	***14.1**. Putting on your socks or tights without help or aids (e.g. sock aid):* 0= Easy 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9

		11= Mahdotonta 99= don't know	10= 10 11= Impossible 99= don't know
Q48	char(2)	***K48**. Kumartuminen vyötäröstä eteenpäin nostaaksenne kynän lattialta ilman...* 0= Helppoa 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Mahdotonta 99= don't know	***14.2**. Bending forward from the waist to pick up a pen from the floor without an aid:* 0= Easy 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Impossible 99= don't know
Q49	char(2)	***K49**. Kurkottaminen korkealle hyllylle ilman apua tai apuvälineitä?* 0= Helppoa 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Mahdotonta 99= don't know	***14.3**. Reaching up to a high shelf without help or aids (e.g. helping hand):* 0= Easy 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Impossible 99= don't know
Q50	char(2)	***K50**. Nouseminen ylös käsinojattomasta tuolista käyttämättä käsiä tai muuta apua?* 0= Helppoa 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8	***14.4**. Getting up out of an armless dining room chair without using your hands or any other help:* 0= Easy 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8

		9= 9 10= 10 11= Mahdotonta 99= don't know	9= 9 10= 10 11= Impossible 99= don't know
Q51	char(2)	***K51**. Nouseminen ylös selinmakuulta lattalta ilman apua?* 0= Helppoa 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Mahdotonta 99= don't know	***14.5**. Getting up off the floor without help from lying on your back:* 0= Easy 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Impossible 99= don't know
Q52	char(2)	***K52**. Seisominen ilman tukea 10 minuuttia vaikeuksitta?* 0= Helppoa 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Mahdotonta 99= don't know	***14.6**. Standing unsupported for 10 minutes without discomfort:* 0= Easy 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Impossible 99= don't know
Q53	char(2)	***K53**. Kiipeämäinen 12-15 askelta käyttämättä kaidetta tai kävelyapuvälinettä?* 0= Helppoa 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8	*14.7. Climbing 12-15 steps without using a handrail or walking aid. **One foot on each step**:* 0= Easy 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7

		9= 9 10= 10 11= Mahdotonta 99= don't know	8= 8 9= 9 10= 10 11= Impossible 99= don't know
Q54	char(2)	***K54**. Katsominen yli olan kääntämättä kehoanne?* 0= Helppoa 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Mahdotonta 99= don't know	***14.8**. Looking over your shoulder without turning your body:* 0= Easy 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Impossible 99= don't know
Q55	char(2)	***K55**. Ruumiillisesti rasittavien tehtävien tekeminen?* 0= Helppoa 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Mahdotonta 99= don't know	***14.9**. Doing physically demanding activities (e.g. physiotherapy exercises, gardening or sports):* 0= Easy 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Impossible 99= don't know
Q56	char(2)	***K56**. Päivittäisten toimintojen tekeminen, joko kotona tai tössä?* 0= Helppoa 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7	***14.10**. Doing a full day's activities whether it be at home or at work:* 0= Easy 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6

		8= 8 9= 9 10= 10 11= Mahdotonta 99= don't know	7= 7 8= 8 9= 9 10= 10 11= Impossible 99= don't know
Q57	char(1)	***K57**. Oletteko kaksonen?* 1= yes 2= no 9= don't know	***15.a**. Are you a **twin**?* 1= yes 2= no 9= don't know
Q58	char(1)	***K58**. Mikäli vastasitte Kyllä, oletteko* *1= identtinen kaksonen* *2= ei-identtinen kaksonen* 9= don't know	***15.a.i**. If YES, are you:* *1= identical twin* *2= non-identical twin* 9= don't know
Q59	char(1)	***K59**. Onko muita kaksosia suvussanne?* 1= yes 2= no 9= don't know	***15.b**. Are there any (other) twins in your family?* 1= yes 2= no 9= don't know
Q60	char(200)	***K60**. Mikäli vastasitte Kyllä, niin määritelkää heidän sukulaisuussuhteenne Teihin:*	***15.b.i**. If YES, please state their relationship to you:*
Q61	char(1)	***K61**. Onko Teillä veljiä tai sisaruksia?* 1= yes 2= no 9= don't know	***16**. Do you have any brothers or sisters?* 1= yes 2= no 9= don't know Note: information on brothers and sisters is in table **relatives**
Q63	char(1)	***K63**. Onko Teillä lapsia?* 1= yes 2= no 9= don't know	***17**. Do you have any children?* 1= yes 2= no 9= don't know Note: information on childrens is in table **relatives**
Q65	char(1)	***K65**. Onko Teidän tietääksenne kenelläkään sukulaisistanne...* 1= yes 2= no 9= don't know	***18.a**. To your knowledge, do any members of your family have ankylosing spondylitis?* 1= yes 2= no 9= don't know
Q66	char(200)	***K66**. Mikäli vastasitte Kyllä, niin*	***18.a.i**. If YES, please*

		mainitkaa kenellä sukulaisella	*state which relative(s):* Note: information on relatives with AS is in table **relatives_srr**
PLACE**	char(50)	*Place of filling this Questionnaire*	*Place of filling this Questionnaire*
DATEFILL	char(10) dd.mm.yyyy	*Date of filling this Questionnaire*	*Date of filling this Questionnaire*
SIGNED	char(1)	*Merkintä, onko allekirjoitus:* 1= yes 2= no 3= puolesta 9= don't know	*Mark if signed personally:* 1= yes 2= no 3= on behalf of 9= don't know

Table RELATIVES
K62-K64.Veljiä, sisaruksia, lapsia, vanhemmat
16., 17. Brothers, sisters, children, parents

Variable	Format	Kuvaus	Description
RECID**	integer	*Record ID, for internal use*	*Record ID, for internal use*
CENTERID	char(2)	*Center ID*: 01-99	*Center ID*: 01-99
FAMILYID	char(4)	*Family ID*: 0001-9999	*Family ID*: 0001-9999
INDIVID	char(3)	*Individual ID:* 000= proband 001-999= other relatives	*Individual ID:* 000= proband 001-999= other relatives
PTID**	char(9)	*for internal use*	*for internal use*
REL_ID	char(7)	*for internal use*	*for internal use*
RELATION	char(2)	Relation: 1= brother 2= sister 3= child 4= father 5= mother 6= father's father 7= father's mother 8= mother's father 9= mother's mother 10= other 99= don't know	Relation: 1= brother 2= sister 3= child 4= father 5= mother 6= father's father 7= father's mother 8= mother's father 9= mother's mother 10= other 99= don't know
L_NAME**	char(30)	*Last name*	*Last name*
F_NAME**	char(50)	*First names(s)*	*First names(s)*
SEX	char(1)	*Sex:* 1= male 2= female 9= don't know	*Sex:* 1= male 2= female 9= don't know

ALIVE	char(1)	*Alive status:* 1= alive 2= dead 9= don't know	*Alive status:* 1= alive 2= dead 9= don't know
PARTIC	char(1)	*Participation:* 1= Osallistuu 2= ei-osallistuu 9= don't know	*Participation:* 1= Osallistuu 2= ei-osallistuu 9= don't know
DBIRTH	char(10) dd.mm.yyyy	*Date of birth*	*Date of birth*
ADDRESS**	50	*Address, street*	*Address, street*
CITY**	30	*Address, city*	*Address, city*
ZIPCODE**	6	*Address, postal code*	*Address, postal code*
TEL_1**	20	*Telephone - 1*	*Telephone - 1*
TEL_2**	20	*Telephone - 2*	*Telephone - 2*

Table RELATIVES_SRR
Tietoja sukulaisista, joilla on selkärankareuma
18.a.i. Other relatives with ankylosing spondylitis

Variable	Format	Kuvaus	Description
RECID**	integer	*Record ID, for internal use*	*Record ID, for internal use*
CENTERID	char(2)	*Center ID*: 01-99	*Center ID*: 01-99
FAMILYID	char(4)	*Family ID*: 0001-9999	*Family ID*: 0001-9999
INDIVID	char(3)	*Individual ID:* 000= proband 001-999= other relatives	*Individual ID:* 000= proband 001-999= other relatives
PTID**	char(9)	*for internal use*	*for internal use*
REL_ID**	char(7)	*for internal use*	*for internal use*
RELATION	char(2)	Relation: 1= brother 2= sister 3= child 4= father 5= mother 6= father's father 7= father's mother 8= mother's father 9= mother's mother 10= other 99= don't know	Relation: 1= brother 2= sister 3= child 4= father 5= mother 6= father's father 7= father's mother 8= mother's father 9= mother's mother 10= other 99= don't know
L_NAME**	char(30)	*Last name*	*Last name*
F_NAME**	char(50)	*First names(s)*	*First names(s)*
SEX	char(1)	*Sex:* 1= male	*Sex:* 1= male

		2= female 9= don't know	2= female 9= don't know
ALIVE	char(1)	*Alive status:* 1= alive 2= dead 9= don't know	*Alive status:* 1= alive 2= dead 9= don't know
PARTIC	char(1)	*Participation:* 1= Osallistuu 2= ei-osallistuu 9= don't know	*Participation:* 1= Osallistuu 2= ei-osallistuu 9= don't know
DBIRTH**	char(10) dd.mm.yyyy	*Date of birth*	*Date of birth*
ADDRESS**	50	*Address, street*	*Address, street*
CITY**	30	*Address, city*	*Address, city*
ZIPCODE**	6	*Address, postal code*	*Address, postal code*
TEL_1**	20	*Telephone - 1*	*Telephone - 1*
TEL_2**	20	*Telephone - 2*	*Telephone - 2*

Table SURGERIES
K33. Mikäli vastasitte KYLLÄ, kertokaa leikkaustyyppi ja leikkausvuosi
9b. Have you ever had surgery for your joints or spine as a result of your AS? If YES, please specify the type of surgey and give year(s)

Variable	Format	Kuvaus	Description
RECID**	integer	*Record ID, for internal use*	*Record ID, for internal use*
CENTERID	char(2)	*Center ID*: 01-99	*Center ID*: 01-99
FAMILYID	char(4)	*Family ID*: 0001-9999	*Family ID*: 0001-9999
INDIVID	char(3)	*Individual ID:* 000= proband 001-999= other relatives	*Individual ID:* 000= proband 001-999= other relatives
PTID**	char(10)	*for internal use*	*for internal use*
YEAR	char(4)	*Year when surgery was performed*	*Year when surgery was performed*
TYPE	char(50	*Type of surgery*	*Type of surgery*

Table MEDICATIONS

K39. Mikäli vastasitte KYLLÄ, ilmoittakaa lääkkeen nimi, annos ja montako kertaa päivässä otatte annoksen

12.a. Are you currently taken any medication for your AS? If YES, please give the name and dosage that is on the bottle or package.

Variable	Format	Kuvaus	Description
RECID**	integer	*Record ID, for internal use*	*Record ID, for internal use*
CENTERID	char(2)	*Center ID*: 01-99	*Center ID*: 01-99
FAMILYID	char(4)	*Family ID*: 0001-9999	*Family ID*: 0001-9999
INDIVID	char(3)	*Individual ID:* 000= proband 001-999= other relatives	*Individual ID:* 000= proband 001-999= other relatives
PTID**	char(9)	*for internal use*	*for internal use*
NAME	char(30)	*Name of medication*	*Name of medication*
DOSE	char(50)	*Dose*	*Dose*
YEAR	char(4)	*Year*	*Year*
EFFECTIV	char(2)	***K40**. Ympyröikää alla olevassa asteikossa se numero, joka parhaiten osoitta lääkityksen tehokkuutta helpottamaan oireitanne:* 0= Ei vaikutusta 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Hyvin tehokas 99= don't know	***12.c**. Place a vertical mark on the line below to indicate the effectiveness of the medication in relieving your symptoms:* 0= No effect 1= 1 2= 2 3= 3 4= 4 5= 5 6= 6 7= 7 8= 8 9= 9 10= 10 11= Very effective 99= don't know

APPENDIX 2

Lysis Buffer

109.4 g of Sucrose

1000 ml of distilled water

1 ml of 1M Tris-HCl (pH 7.8)

1.2 ml 4.9M $MgCl_2$

10 ml of Triton X-100

6M Guanidinium hydrochloride GuHCl

286.59 g solid guanidinium hydrochloride

500 ml distilled water

Autoclave at 121°C for 20 mins

7.5M Ammonium Acetate

57.81 g solid NH_4Ac

100 ml distilled water

Autoclave at 121°C for 20 mins

10% Sodium Dodecyl Sulphate (SDS)

10 g SDS power

100 ml distilled water

20 mg/ml Proteinase K

1 g Proteinase K

50 ml distilled water

1xTE buffer

0.01M Tris-HCl

0.1M EDTA

pH = 8.0

1 x TBE solution

0.1M Tris

0.09M Boric Acid

0.001M EDTA

pH = 8.0

APPENDIX 3

Alleles / haplotypes / microsatellites recognised	Orientation	Primer Sequences 5' - 3'	Product length (bp)	$MgCl_2$ concentration
TGFB1 -800 G to -509 C	Sense Antisense	AGGGACTCTGCCTCCAACG GGGCAACAGGACACCTGAG	291	2.5 mM
TGFB1 -800 G to -509 T	Sense Antisense	AGGGACTCTGCCTCCAACG GGGCAACAGGACACCTGAA	291	2.5 mM
TGFB1 -800 A to -509 C	Sense Antisense	AGGGACTCTGCCTCCAACA GGGCAACAGGACACCTGAG	291	2.5 mM
TGFB1 -800 A to -509 T	Sense Antisense	AGGGACTCTGCCTCCAACA GGGCAACAGGACACCTGAA	291	2.5 mM
TGFB1 868 T	Sense Antisense	CCCCAGACCTCGGGCGC GCAGCGGTAGCAGCAGCA	416	1.5 mM
TGFB1 868 C	Sense Antisense	CCCCAGACCTCGGGCGC GCAGCGGTAGCAGCAGCG	416	1.5 mM
TGFB1 913 G	Sense Antisense	CCCCAGACCTCGGGCGC TAGTCCCGCGGCCGGCC	428	1.25 mM
TGFB1 913 C	Sense Antisense	CCCCAGACCTCGGGCGC TAGTCCCGCGGCCGGCG	428	1.25 mM
TGFB1 1627 C	Sense Antisense	GCTTCCCTCTCGCCACTC GGGCCCTCTCCAGCGGGG	200	2.25 mM
TGFB1 1627 T	Sense Antisense	GCTTCCCTCTCGCCACTC GGGCCCTCTCCAGCGGGA	200	2.25 mM
D19S217	Sense Antisense	GGGGTTGATTGAAGTTGGTT CAAGACCCATACCCATGA	219-233	2 mM
D19S223	Sense Antisense	CAAAATCGAGGTGCATAGAA ACCATGACTGGCTAATTGTG	228-246	2 mM
D19S421	Sense Antisense	AGCCCTTGCTGGTTTATG GAGCCTGTGGTGGAGC	188-220	2 mM

The cycling conditions for the *TGFB1* SNPs were as follows: 96°C for 1 min, 5 cycles of 96°C for 25 s, 70°C for 45 s, 72°C for 45 s, 21 cycles of 96°C for 25 s, 65°C for 50 s, 72°C for 45 s and 4 cycles of 96°C for 25 s, 55°C for 1 min, 72°C for 2 min.

The cycling conditions for the *TGFB1* microsatellites were as follows: 32 cycles of denaturation at 94°C for 1 min, annealing 54°C (D19217 and D19223) or 50°C (D19421) for 1 min and extension 72°C for 1 min.